Some results concerning certain solvable directed polymer models and non-linear stochastic partial differential equations

By

Christopher Janjigian

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Mathematics)

at the

UNIVERSITY OF WISCONSIN – MADISON

2016

Date of final oral examination: July 25, 2016

The dissertation is approved by the following members of the Final Oral Committee:

Benedek Valkó, Associate Professor, Mathematics

Timo Seppäläinen, Professor, Mathematics

Jun Yin, Associate Professor, Mathematics

David Anderson, Associate Professor, Mathematics

Jean-Luc Thiffeault, Professor, Mathematics

Abstract

This dissertation is broken into two essentially unrelated chapters. The first chapter considers exact computations of large deviation rate functions in various solvable 1+1 dimensional directed polymer models. The models considered include point-to-point and stationary versions of an inhomogeneous directed last passage percolation model, the O'Connell-Yor polymer, and the Brownian directed percolation model. The work on the inhomogeneous corner growth model is joint with Elnur Emrah. The second chapter deals with particle representations for a class of nonlinear stochastic partial differential equations and is based on joint work with Dan Crisan and Tom Kurtz.

Acknowledgements

I am extremely grateful to the many people whose support have made it possible to complete this dissertation. I would like to begin by thanking my Ph.D. adviser, Benedek Valkó, for his patience, his support, and his friendship. This project would not have been completed without his help. For the same reasons, I am deeply indebted to Tom Kurtz and Timo Seppäläinen, who have both treated me like their own students and have provided me with research problems and extensive feedback. I am especially grateful to Benedek and Tom for supporting me through their NSF grants DMS 1053280 and DMS 1106424. I would also like to thank the rest of the probability group at the University of Wisconsin - Madison for making this department an excellent place to work.

I would like to thank my coauthors Dan Crisan, Elnur Emrah, Tom Kurtz, for their willingness to work with me and for their patience. I have always enjoyed mathematics most when working with others. The projects with Dan, Elnur, Tom were enjoyable to work on in large part due to having such wonderful coauthors to work with.

Most of all I am thankful to my friends and family, without whose support I would never have made it this far.

List of Figures

Contents

Chapter 1

Introduction and overview

This thesis consists of two main chapters, which study essentially disjoint topics. Both chapters include more detailed introductions to the questions considered here, but we will begin by giving a brief overview of the results obtained in this dissertation.

Chapter 2 considers results on large deviations of the free energy in certain directed polymer models. Formally, the model of a directed polymer in a random environment is a measure on paths. To construct such a model on the lattice $\mathbb{Z}^d \times \mathbb{Z}_+$, we assign weights $\{W(\mathbf{x},t)\}_{(\mathbf{x},t)\in\mathbb{Z}^d \times \mathbb{Z}_+}$ to each site of the lattice. The directed polymer measure is defined as a perturbation on a reference measure: for concreteness, let P denote the law of a simple symmetric random walk. Fixing a parameter $\beta > 0$ (the inverse temperature) and a polymer length $T \in \mathbb{Z}_+$, the measure on paths is given by

$$\mu_T(\pi) = Z_T(\beta)^{-1} e^{\beta \sum_{t=1}^T W(\pi(t),t)} P_T^{\mathrm{ref}}(\pi).$$

See Chapter 2, Subsection 2.1.1 for a more detailed and precise introduction to the model. Our interest is in large deviation properties of the free energy $T^{-1}\log Z_T(\beta)$ in the case $d = 1$ for certain choices of the weight distributions. Large deviation theory aims to understand precise exponential asymptotics for the probability of rare events. The regime studied here corresponds to the case when the free energy is unusually large. In the case of the directed polymer models we study, these unusually large values of the free energy are connected to the physical phenomenon of intermittency. Some connections to the relevant literature and a discussion of this connection are included in Chapter 2, Subsection 2.1.3.

The precise models studied here are discussed in detail in Chapter 2, Subsection 2.2. They consist of a small collection of 'exactly solvable'—that is, models for which explicit computation is possible—and include an inhomogeneous generalization of the classical exponential directed last passage percolation model, the O'Connell-Yor polymer model, and Brownian directed last passage percolation. The primary results for each of these models are computations of large deviation rate functions and moment Lyapunov exponents corresponding to right tail large deviations in the point-to-point versions of the polymer model (i.e. where P is the law of a random bridge). Additionally, in each case some results are obtained for stationary versions of the model. The results for the inhomogeneous generalization of exponential last passage percolation appear in [32]. Most of the results on the O'Connell-Yor polymer appear in [49], though some additional results are presented here. Similarly, most of the results for Brownian directed percolation appear in [50].

Chapter 3 concerns particle representations for a class of non-linear stochastic partial differential equations with multiplicative noise and Dirichlet boundary conditions. This work is complementary to the results of the paper [28], which is joint work with Dan Crisan and Tom Kurtz. In that paper, a similar particle representation was obtained for a different class of stochastic partial differential equations with additive noise and Dirichlet boundary conditions. Particle representations and approximations of this type originate in the study of the McKean-Vlasov problem and appear for example in the theory and practice of non-linear filtering. A simple example covered by the results of [28] is the stochastic Allen-Cahn equation with time-white space-colored noise forcing,

$$\partial_t u = \Delta u + u - u^3 + \xi$$

on a domain D, subject to the boundary condition that (in a certain sense) $u(t,x) = g(x)$ on ∂D.

Both the results presented here and the results in [28] are extensions of the work of Kurtz and Xiong in [63] to stochastic partial differential equations on domains. Here, as in [63], the non-linearity in the stochastic partial differential equation is a bounded and Lipschitz continuous functional of the de Finetti measure which serves as a representation of the solution to the stochastic partial differential equation. This differs from [28], where the non-linearity is a Lipschitz function of the density of the solution. The latter is suited for equation of the type shown above, while the former sometimes appears in the context of filtering.

Chapter 2

Large deviations of the free energy in certain solvable directed polymer models

2.1 Introduction

2.1.1 Directed polymers in random environments

The model of a directed polymer in a random environment was introduced in the physics literature in [46] in order to model the domain wall in a ferromagnetic Ising model with random impurities. Shortly thereafter, it was observed [47, 54] that in the two dimensional model the numerically observed scaling exponents for the transverse fluctuations of the domain wall interface and the pinning energies also appeared numerically and theoretically in other related contexts [35, 88]. These works suggested the presence of some universal features in certain 1+1 dimensional models of interfaces roughened by impurities. Early mathematical work on the model in dimensions 3+1 and higher followed in [14, 48]. The directed polymer model itself, these characteristic scaling exponents in 1+1 dimensions, and other universal aspects of models of this type have since appeared in a wide range of physical and mathematical situations. See for example [43, 42, 59] for physical surveys and [23, 25, 30, 75, 76] for mathematical surveys. We begin with an introduction to the model and some comments about the specific questions studied in this dissertation.

The directed polymer model is formally a probability measure on paths in a disordered (random) environment. The name comes from the interpretation of a random path drawn from this measure as describing the shape of a polymer chain. Consider the lattice $\mathbb{Z}^d \times \mathbb{Z}_+$ and let $W(\mathbf{x}, t)$ be a family of real valued (random) weights indexed by $(\mathbf{x}, t) \in \mathbb{Z}^d \times \mathbb{Z}_+$. Let $P^{\text{ref}}(\cdot)$ denote the law of a random walk on $\mathbb{Z}^d$ and $P_T^{\text{ref}}(\cdot)$ denote the restriction of P^{ref} to times $t \in \{0, \ldots, T\}$. Fix a parameter $\beta \geqslant 0$, which we can interpret as an inverse temperature. The polymer measure is

$$\mu_T(\pi) = Z_T(\beta)^{-1} e^{\beta \sum_{t=1}^{T} W(\pi(t),t)} P_T^{\text{ref}}(\pi), \tag{2.1.1}$$

where Z is a normalizing constant referred to as the partition function. We take the convention common in the mathematical literature of making the exponent positive. Paths for which $\sum_{t=1}^{T} W(\pi(t), t)$ is large are then assigned greater weight by the measure μ_T. It should be mentioned that unless $\beta = 0$, this family of measures is in general *not* consistent as T varies. That is, if $T_1 < T_2$ then integrating out the distribution of $(\pi(T_1+1), \ldots, \pi(T_2))$ from μ_{T_2} does not result in the measure μ_{T_1}.

Informally, sites with $W(\mathbf{x}, t) > 0$ can be viewed as favorable to the polymer chain, while sites with $W(\mathbf{x}, t) < 0$ are unfavorable. A simple physical picture to keep in mind (taken from [23]) would be to imagine a hydrophilic polymer chain floating in water and to consider the case that $W(\mathbf{x}, t) \in \{-1, 1\}$. The lattice here can be viewed as representing the sites where monomers can be located while nearest-neighbor edges between these vertices can be viewed as possible locations for chemical bonds. We can interpret the sites with $W(\mathbf{x}, t) = -1$ as sites with hydrophobic impurities and sites with $W(\mathbf{x}, t) = 1$ as those without. For β sufficiently large, one can see that the typical shape of the chain π under μ_T will tend to be one where $\pi(t) = \mathbf{x}$ for a large number of sites $(\mathbf{x}, t)$ with $W(\mathbf{x}, t) = 1$.

This measure should be thought of as a model of the shape of a polymer at thermal equilibrium with a fixed realization of this environment. See for example the discussions in [23, 30, 43]. The model considered here is not the most general random polymer measure considered in the literature and indeed lacks some physically interesting features if one would like to model actual polymers. In particular, having chosen to make the polymer 'directed' by requiring that the second coordinate (time) increase in each step, we lose self-interactions of the chain. This cost is somewhat compensated by the fact that the model becomes considerably more tractable with this choice.

As a concrete example, consider the case that P^{ref} is the law of a simple symmetric random walk π on $\mathbb{Z}$ and let P_T^{ref} be the restriction of this law to times $t \in \{0, 1, \ldots, T\}$. There are two natural graphical views of the polymer paths that appear in the literature: one with the coordinates given by $(x, t) \in \mathbb{Z} \times \mathbb{Z}_+$ and a rotated picture with coordinates $(i, j) \in \mathbb{Z}_+ \times \mathbb{Z}_+$. As the name suggests, the rotated picture is obtained from the space-time picture with a 45 degree rotation. The coordinates (i, j) are related to the coordinates (x, t) by $i + 1 - 1 = t$ and $i - j = x$. See Figure 1 and note that in general we will only draw the sites in the lattice which can be reached by the paths of the walk with positive probability.

The previous model, where the reference measure P_T^{ref} is the law of a simple symmetric random walk, is often referred to as the point-to-line polymer. It is natural to consider the model where this reference measure is replaced by that of a random bridge. Indeed, this is the model that is the object of interest in what follows. Formally, fix $\mathbf{y} \in \mathbb{Z}^d$ and let $P_{\mathbf{y},T}^{\mathrm{ref}}(\pi)$ denote the law of a random bridge $\pi(t)$ on $\mathbb{Z}^d$ with $\pi(0) = \mathbf{0}$ and $\pi(T) = \mathbf{y}$. In this case, the polymer measure is given by

$$\mu_{\mathbf{y},T}(\pi) = Z_{\mathbf{y},T}(\beta)^{-1} e^{\beta \sum_{t=1}^{T} W(\pi(t),t)} dP_{\mathbf{y},T}^{\mathrm{ref}}(\pi). \tag{2.1.2}$$

Once again, in the case $d = 1$ it is natural to work in the rotated picture with coordinates (i, j) rather than (x, t). See Figure 2.

In both cases, it is also be natural to consider the "zero temperature" polymer model,

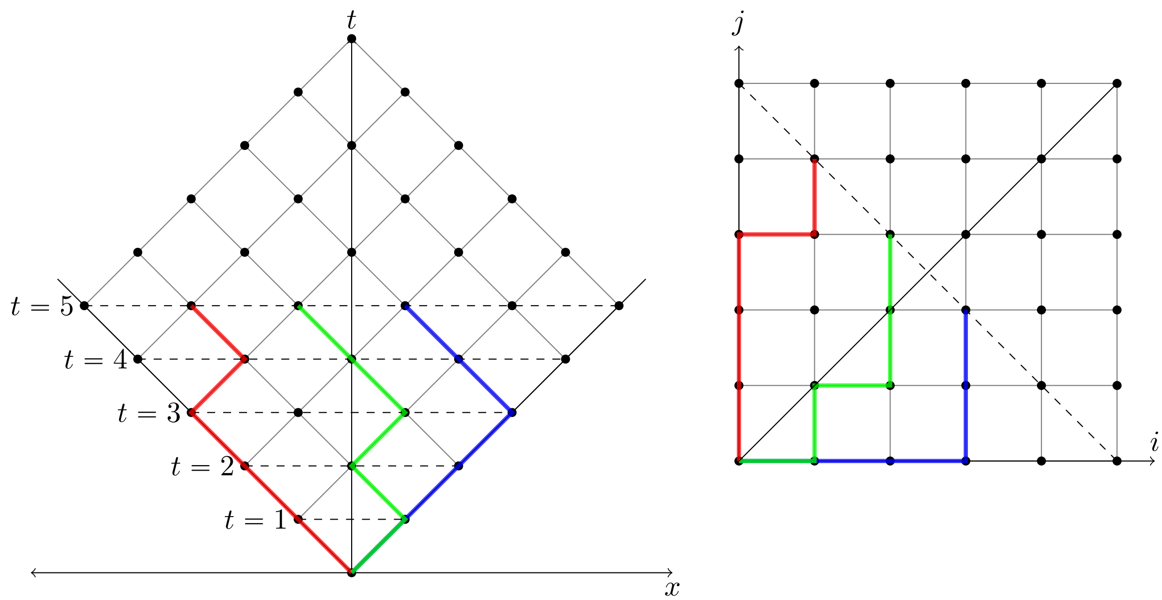


Figure 1: Three paths of a simple random walk π on $\mathbb{Z}$ with $\pi(0) = 0$ in the space-time picture and the rotated picture up to time $t = 5$. Time runs along the main diagonal in the rotated picture.

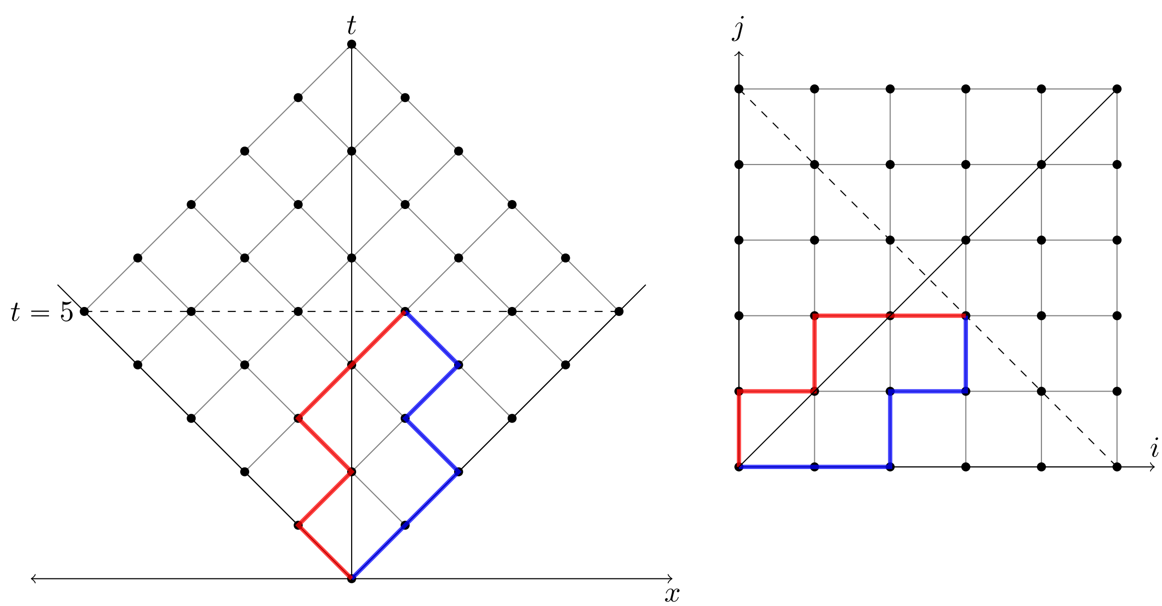


Figure 2: Two paths of a simple random bridge π on $\mathbb{Z}$ in the space-time picture and the rotated picture with $\pi(0) = 0$ and $\pi(5) = 1$.

which is given by taking the limit $\beta \to \infty$ in the previous expressions. The terminology comes from the interpretation of β as being a (multiplicative) inverse temperature, as is common in the statistical physics literature. Define

$$\lim_{\beta\to\infty} \beta^{-1} \log Z_T(\beta) := L(T) = \max_{\pi} \sum_{t=1}^{T} W(\pi(t), t)$$
$$\lim_{\beta\to\infty} \beta^{-1} \log Z_{x,T}(\beta) := L(x,T) = \max_{\pi} \sum_{t=1}^{T} W(\pi(t), t)$$

where the maxima run over paths π which are supported by the reference measure P_T^{ref} or $P_{\mathbf{y},T}^{\mathrm{ref}}$ respectively. This zero temperature model is typically referred to as directed last passage percolation and the maximum over paths in the previous expressions are referred to as last passage times.

2.1.2 Free energy fluctuations and the KPZ class in $d = 1$

Under fairly mild conditions on the weights $W(\mathbf{x}, t)$ (see for example [89]), one can show that the limits

$$\rho_{pl} = \lim_{n\to\infty} n^{-1} \log Z_n(1), \qquad \rho_{pp}(s) = \lim_{n\to\infty} n^{-1} \log Z_{\lfloor ns \rfloor, n}(1)$$
$$g_{pl} = \lim_{n\to\infty} n^{-1} L(n), \qquad g_{pp}(s) = \lim_{n\to\infty} n^{-1} L(\lfloor ns \rfloor, n)$$

exist almost surely. The terms in the first line are typically referred to as free energies, while the terms in the second are often referred to as time constants. Note that we have suppressed the dependence on β in the first line. In dimensions 3+1 and higher, one sees a phase transition in the behavior of the polymer model as β varies [23]. We consider the 1+1 dimensional model here and will omit the dependence on β by fixing $\beta = 1$.

One of the numerical observation in [46] was that the fluctuations of $\log Z_n(1)$ about $n\rho_{pl}$ in the 1+1 dimensional case should scale as $n^{1/3}$. There has been enormous recent progress on understanding the limiting distributions under this scaling. These limits are expected to be universal for a wide class of interacting particle systems, growth models, and directed polymer models, but the exact distribution is expected to depend on the initial (or terminal) conditions of the model. For example, in the point-to-point case, it is widely expected that the following conjecture holds (see for example [12, 25, 85]).

Conjecture 2.1.1. *When $d = 1$, for a wide class of distributions on the i.i.d. weights $\{W(x,t)\}_{x,t}$, there exists a constant c depending on the distribution of $W(1,1)$ so that*

$$\lim_{n\to\infty} P\left(\frac{\log Z_{\lfloor ns \rfloor,n}(1) - n\rho_{pp}(s)}{cn^{1/3}} \leqslant r\right) = F_2(r)$$

where $F_2(r)$ is the cumulative distribution function of the Tracy-Widom GUE distribution.

The same conjecture is expected to hold with $\log Z_{\lfloor ns \rfloor,n}(1)$ replaced with $L(\lfloor ns \rfloor, n)$ and $\rho_{pp}(s)$ replaced with $g_{pp}(s)$. Although there has not been much progress toward true universality, this conjecture has been checked for certain solvable models.

The previous conjecture needs to be modified if one changes the paths in the polymer model. For example, in the point-to-line polymer model the Tracy-Widom GUE distribution should be replaced by the Tracy-Widom GOE distribution. A more general description of the conjectured limiting distributions can be seen in [25, Figure 4]. There is a process level version of this conjecture though even less is known rigorously; see [26].

The class of models for which appropriate versions and generalizations of this conjecture are expected to hold is known as the Kardar-Parisi-Zhang (KPZ) universality class. This class has attracted substantial research interest in the last two decades. See for example the surveys [25, 42, 75, 76]. The namesake of the class is the Kardar-Parisi-Zhang (KPZ) equation, which describes the evolution of the free energy of the continuum directed polymer [1, 2]. To define this process, it is helpful to start by considering the stochastic heat equation with multiplicative noise, which describes the partition function of the continuum directed polymer:

$$\partial_t Z = \frac{1}{2}\partial_x^2 Z + Z\dot{W}. \tag{2.1.3}$$

Here $\dot{W}$ is space-time white noise on $\mathbb{R} \times \mathbb{R}_+$ and for sufficiently nice initial data solutions to this equation are typically understood to be mild solutions in the sense of Walsh [90]. The initial conditions $Z(x,0) = \delta(x)$ and $Z(x,0) = 1$ correspond to the point-to-point and point-to-line models respectively. A formal computation assuming that $\dot{W}$ was a smooth function shows that $h = \log Z$ should solve the Kardar-Parisi-Zhang equation

$$\partial_t h = \frac{1}{2}\partial_x^2 h + \frac{1}{2}(\partial_x h)^2 + \dot{W}. \tag{2.1.4}$$

We take $h = \log Z$ for Z solving (2.1.3) to be the definition of a solution to (2.1.4). This was shown to be the physically correct notion of a solution by Bertini and Giacomin in 1997 [10]. A direct definition of an appropriately renormalized solution to (2.1.4) (on $\mathbb{T} \times \mathbb{R}_+$, rather than $\mathbb{R} \times \mathbb{R}_+$) came in 2011 with the Fields Medal winning work of Hairer on the theory of regularity structures [40, 41]. Recently, an alternative approach in the same setting was proposed by Gubinelli and Perkowski [39], using the language of paracontrolled distributions.

The KPZ equation itself was recently shown to lie in the universality class [3, Corollary 1.7], in that with $Z(x,0) = \delta(x)$, for each $x \in \mathbb{R}$, $2^{1/3}t^{-1/3}(\log Z(xt^{2/3}, t) - t/24$ converges to the Tracy-Widom GUE distribution as $t \to \infty$.

2.1.3 Free energy large deviations and annealed moment Lyapunov exponents

Large deviation theory

Large deviation theory is a branch of probability theory studying sequences of events with exponentially small probabilities. All of the random variables considered here will take values in $\mathbb{R}$, so we state the definition of a large deviation principle at this level of generality.

Definition 2.1.2. *Given a sequence of real-valued random variables $\{X_n\}$, we say that the distributions of $\{X_n\}$ satisfy a large deviation principle with rate r_n and good rate function $I(\cdot)$*

if $I(\cdot)$ is lower semi-continuous, has compact sub-level sets, and for all Borel sets B

$$\inf_{x\in\overline{B}} I(x) \leqslant \underline{\lim}_{n\to\infty} -r_n^{-1}\log P\left(X_n \in \overline{B}\right) \leqslant \overline{\lim}_{n\to\infty} -r_n^{-1}\log P\left(X_n \in B^o\right) \leqslant \inf_{x\in B^o} I(x),$$

where B^o denotes the interior of B and $\overline{B}$ denotes the closure.

Two classical results in large deviation theory connect these large deviation rate functions to exponential moments of (functions of) the sequence of random variables. There is a result due to Varadhan which shows that if a large deviation principle holds then under a moment assumption one can recover asymptotics of exponential moments of the random variables.

Lemma 2.1.3 (Varadhan's Lemma, [29] Theorem 4.3.1)**.** *Let $\phi : \mathbb{R} \to \mathbb{R}$ be continuous and suppose that the distributions of $\{X_n\}$ satisfy a large deviation principle with rate r_n and good rate function $I(\cdot)$. Suppose further that for some $\gamma > 1$,*

$$\overline{\lim_{n\to\infty}}\, r_n^{-1}\log E\left[e^{r_n\gamma\phi(X_n)}\right] < \infty.$$

Then

$$\overline{\lim_{n\to\infty}}\, r_n^{-1}\log E\left[e^{r_n\phi(X_n)}\right] = \sup_{x\in\mathbb{R}}\{\phi(x) - I(x)\}.$$

There is also a partial converse due to Gärtner and Ellis, which says that under smoothness conditions the reverse also holds.

Theorem (Gärtner-Ellis Theorem, [29] Theorem 2.3.6)**.** *Suppose that for $\lambda \in \mathbb{R}$, the limit*

$$\Lambda(\lambda) := \lim_{n\to\infty} r_n^{-1}\log E\left[e^{\lambda r_n X_n}\right]$$

exists as an extended real number in $(-\infty, \infty]$ and that the function $\lambda \mapsto \Lambda(\lambda)$ is lower semi-continuous. Let $D_\Lambda = \{\lambda : \Lambda(\lambda) < \infty\}$ and suppose that its interior, D_Λ^o, is non-empty, $0 \in D_\Lambda^o$, and that $\Lambda(\cdot)$ is a differentiable function on D_Λ^o. If in addition Λ is steep *in the sense that $\Lambda'(\lambda_n) \to \infty$ whenever λ_n is a sequence in D_Λ^o converging to a boundary point of D_Λ^o, then the distribution of X_n satisfies a large deviation principle with rate r_n and rate function $\Lambda^*(r) = \sup_{\lambda\in\mathbb{R}}\{\lambda r - \Lambda(\lambda)\}$.*

Corollary 2.1.4. *Suppose that for $\lambda \in \mathbb{R}$, the limit*

$$\Lambda(\lambda) := \lim_{n\to\infty} r_n^{-1} \log E\left[e^{\lambda r_n X_n}\right]$$

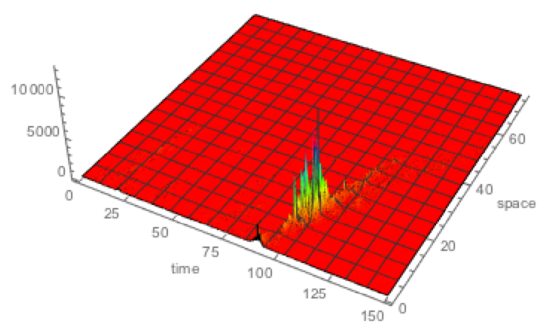
exists as a real number and that $\Lambda(\cdot)$ is a differentiable function on $\mathbb{R}$. Then the distribution of X_n satisfies a large deviation principle with rate r_n and rate function $\Lambda^*(r) = \sup_{\lambda\in\mathbb{R}}\{\lambda r - \Lambda(\lambda)\}$.

Large deviations of the free energy and Lyapunov exponents

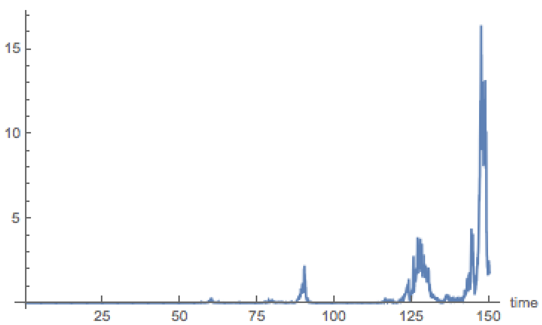
There are two interesting regimes for large deviations of the free energy in point-to-point directed polymers with up-right paths (as in the second frame of Figure 2) when $d = 1$ and the weights $\{W(i,j)\}$ are i.i.d.. The first is the regime we study, which corresponds to right-tail large deviations–meaning the regime in which the free energy is unusually large. Heuristically, one can guess that the correct rate for such large deviations should be $r_n = n$ by viewing the partition function as an integral over paths. The partition function can become unusually large if a single path is unusually large. Since there are $O(n)$ sites on a path from $(1,1)$ to $(\lfloor ns \rfloor, n)$ and the environment is i.i.d., under mild assumptions direct computation shows that the large deviations for a single path will occur with rate $r_n = n$. In contrast, because the weights are always positive, a large deviation in which the free energy is unusually small constrains all paths. One might guess that this imposes a constraint on the $O(n^2)$ weights that influence admissible paths from $(1,1)$ to $(\lfloor ns \rfloor, n)$ and so the rate should be $r_n = n^2$. These large deviations are more complicated than the right tail large deviations and one can show that the rate is not quite universal; see [8, 27].

When $X_n = n^{-1} \log Z_{\lfloor ns \rfloor, n}(1)$ or $X_n = n^{-1} L_{\lfloor ns \rfloor, n}$ and $r_n = n$, we will refer to the values of the exponential moments appearing in the Gärtner-Ellis theorem as (annealed moment) Lyapunov exponents. Computation and estimation of Lyapunov exponents for various generalizations of (2.1.3) have attracted attention in recent years in connection with the phenomenon of intermittency. See for example [16, 24, 56]. Although this is not the focus of this thesis, we briefly review this connection.

Physical intermittency is the tendency of a field to exhibit extreme clumping, meaning that the mass of the field is concentrated in a collection of small regions which are separated by large voids. As will be discussed in Section 2.2.1 below, the partition function in one of the models studied in what follows (the O'Connell-Yor model) can be viewed as a spatial discretization of (2.1.3). The precise definition of this partition function is given in (2.2.2). Figure 3 shows a single simulation of the partition function as a field and along a single spatial line in this model to illustrate the phenomenon. As is suggested by Figure 3, typical values of the (normalized)



(a) $e^{-\frac{3}{2}t}Z_{0,k}(0,t)$ for $k \leqslant 75, t \leqslant 150$



(b) $e^{-\frac{3}{2}t}Z_{0,75}(0,t)$, $t \leqslant 150$

Figure 3: A simulation of the normalized partition function in the O'Connell-Yor polymer.

partition function are small. Indeed from the free energy limit for this model, one can see that the typical values of the partition function for large enough n should be less than one, because the normalized limiting free energy is strictly negative. See for example Lemma 2.2.1 below. The largest peaks in this simulation are on the order 10^4, while the largest peaks along the top line are on the order 10^1. Repeating simulations of this type suggests that the model is typically intermittent in the sense that the largest peaks contain most of the mass (measured by taking space-time averages) of the field. It is shown below that this partition function satisfies the mathematical definition of intermittency, which we now introduce with an example. This was also previously shown in [16].

A concrete theorem proving mathematical intermittency in the case of (2.1.3) was recently proven in the case that $Z(x, 0) = \delta_0(x)$ in [16].

Theorem ([16], Appendix A.2)**.** *Suppose $Z(x, t)$ is the mild solution to (2.1.3) with initial data $Z(x, 0) = \delta_0(x)$. Then for all $\lambda \in \mathbb{N}$,*

$$\lim_{t \to \infty} \frac{1}{t} \log E\left[Z(0, t)^\lambda\right] = \frac{\lambda^3 - \lambda}{24}.$$

A similar result for other initial data can be found in [11, 21]. Mathematical intermittency, as described for example in the book of Carmona and Molchanov [20], is defined by the condition that the function $\lambda \mapsto \lambda^{-1}\Lambda(\lambda)$ is strictly increasing on the interval $[2, \infty)$. Under some ergodicity hypotheses, one can show that if this condition holds, then it will imply a separation of scales similar to what is seen in Figure 3.

Our goal is to explicitly compute large deviation rate functions at rate n and to obtain the associated moment Lyapunov exponents. For the reasons discussed above, in the positive temperature temperature models, which are in some sense discretizations of (2.1.3), results of this type can be viewed as giving some information about intermittency. In both the positive and zero temperature models, computations of these rate functions and Lyapunov exponents provide a more complete picture of the behavior of models in the KPZ class and give some insight into what these models look like when they are not behaving as one might expect based on ideas like Conjecture 2.1.1.

2.2 Models considered and statements of results

This dissertation focuses on a class of models which are "solvable" in the sense that explicit computation of many quantities of interest is possible. This condition is extremely restrictive: only a handful of such models are known and they are only known when $d = 1$. Before turning to the proofs, we begin by introducing the specific models that are studied in this dissertation, recalling some key results about these models, and stating the main results that are proven in what follows.

2.2.1 The O'Connell-Yor polymer

The O'Connell-Yor polymer model was originally introduced in [71] in connection with a generalization of the Brownian queueing model. Based on the work of Matsumoto and Yor [67], O'Connell and Yor were able to show the existence of a stationary version of this model satisfying an analogue of Burke's theorem for M/M/1 queues. This property forms the basis for the computation of the large deviation rate function in Section 2.4. Subsequent work on the representation theoretic underpinnings of the exact solvability of this model can be found in the work of Borodin and Corwin on Macdonald processes [15] and the work of O'Connell connecting this polymer to the quantum Toda lattice [70].

Concretely, it is a semi-discrete model of a directed polymer in a random environment where the random walk paths are given by the sample paths of a Poisson bridge and the random environment is space-time white noise on $\mathbb{R}_+ \times \mathbb{Z}_+$. Let $\{B_i\}_{i=0}^{\infty}$ be a family of independent two-sided standard Brownian motions. For $t \in \mathbb{R}_+$ and $n \in \mathbb{Z}_+$ let $P_{t,n}^{\text{ref}}(\cdot)$ denote the law of a Poisson bridge on $[0, t]$ with $\eta(0) = 0$ and $\eta(t) = n$. In this model the 'energy' of a path is given by

$$H(\eta) = \int_0^t B_{\eta(s)}(ds) = \sum_{i=0}^{n} B_i(t_{i+1}) - B_i(t_i)$$

where $\{t_i\}$ denotes the collection of jumps of $\eta(\cdot)$ on $[s,t]$, with the convention that $t_0 = 0, t_{n+1} = t$. See Figure 4 for an example of a path drawn from $P^{\text{ref}}_{t,7}(\cdot)$.

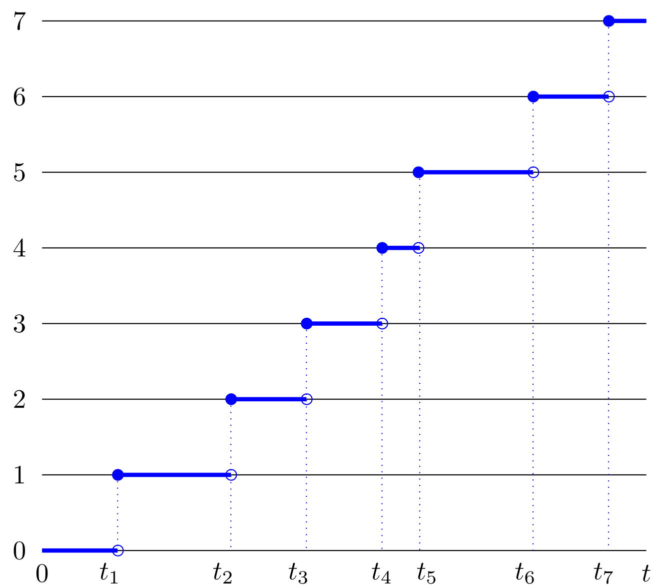


Figure 4: A sample path of a Poisson bridge with $\eta(0) = 0$ and $\eta(t) = 7$ and jumps at t_i

The point to point polymer measure is given by

$$\mu^{\beta}_{t,n}(d\eta) = Z_{t,n}(\beta)^{-1} e^{\beta H(\eta)} P^{\text{ref}}_{t,n}(d\eta).$$

Note that a Poisson bridge η on $[0,t]$ with $\eta(0) = 0$ and $\eta(t) = n$ is uniquely identified by the locations of its n jumps, $\{t_i\}_{i=1}^{n}$, and that these jump are uniformly distributed on the Weyl chamber $A_{n,t} = \{0 < s_1 < \cdots < s_n < t\}$. This model appears with several slightly different definitions of the reference measure in the literature. For example, following [71], we take the convention that $B_i(s,t) = B_i(t) - B_i(s)$ and define

$$Z_n(\beta) = \int\limits_{0<s_1<\cdots<s_{n-1}<n} \exp\left[\beta\left(B_0(0,s_1) + \cdots + B_n(s_{n-1},n)\right)\right] ds_1 \ldots ds_{n-1}. \tag{2.2.1}$$

Computation shows that $Z_{n,n}(\beta) = |A_{n,n}|^{-1} Z_n(\beta)$. We can think of $Z_n(\beta)$ as being the partition function for a polymer measure where we re-weight every path by multiplying by a deterministic constant.

Our primary interest is in the behavior of $Z_n(\beta)$ for large n. In order to highlight various features of this model, we will introduce various other normalizations of the polymer measure and partition function in what follows. We begin by considering the partition functions for more general paths, without defining the associated polymer measures. Let $j, n \in \mathbb{Z}_+$ and $u, t \in \mathbb{R}_+$, where $j < n$ and $u \leqslant t$. Set

$$Z_{j,n}(u,t) = \int\limits_{u<u_j<\cdots<u_{n-1}<t} e^{B_j(u,u_j)+\sum_{i=j+1}^{n-1} B_i(u_{i-1},u_i)+B_n(u_{n-1},t)} du_j \dots du_{n-1}. \tag{2.2.2}$$

For the case $j = n$, we define

$$Z_{j,j}(u,t) = e^{B_j(u,t)}. \tag{2.2.3}$$

We will refer to the j, n variables as space and the u, t variables as time. Translation invariance of Brownian motion and our assumption that the environment is i.i.d. immediately imply that the distribution of these partition function is shift invariant. It follows from Brownian scaling that for $\beta > 0$ and $n > 1$ we have

$$Z_n(\beta) \stackrel{d}{=} \beta^{-2(n-1)} Z_{0,n}(0, \beta^2 n).$$

Distributional results for partition functions of the form $Z_{j,n}(u,t)$ can then be translated into results for $Z_n(\beta)$ using this identity.

The free energy for (2.2.1) was computed in [69]. We have

Lemma 2.2.1 ([69])**.** *Fix $s, t \in (0, \infty)$. Then the almost sure limit*

$$\rho(s,t) = \lim_{n\to\infty} \frac{1}{n} \log Z_{1,\lfloor ns \rfloor}(0, nt)$$

exists and is given by

$$\rho(s,t) = \min_{\theta>0} \{\theta t - s\Psi_0(\theta)\} = t\Psi_1^{-1}\left(\frac{t}{s}\right) - s\Psi_0\left(\Psi_1^{-1}\left(\frac{t}{s}\right)\right).$$

Here, for $\theta > 0$, $\Psi_0(\theta) = \frac{d}{d\theta}\log\Gamma(\theta)$ is the logarithmic derivative of the Gamma function, which is typically referred to as the Digamma function. The polygamma functions are recursively defined for $n \in \mathbb{N}$ by $\Psi_n(\theta) = \frac{d}{d\theta}\Psi_{n-1}(\theta)$. There is a result analogous to Lemma 2.2.1 for the almost sure limit of $n^{-1}Z_n(\beta)$, which is also presented in [69]. The fluctuation result for this model is due to Borodin, Corwin, and Ferrari [17].

Theorem ([17], Theorem 1.3)**.** *Let $t > 0$ and $r \in \mathbb{R}$, then*

$$\lim_{n\to\infty} P\left(\frac{\log Z_{1,n}(0,nt) - n\rho(1,t)}{\left(-\frac{1}{2}\Psi_2\left(\Psi_1^{-1}(t)\right)\right)^{\frac{1}{3}} n^{\frac{1}{3}}} \leqslant r\right) = F_2(r)$$

where $F_2(r)$ is the CDF of the Tracy-Widom GUE distribution.

Before discussing the previous work on large deviations, it is helpful to introduce another normalization of the partition function. Direct computation shows that if we define $X_n(t) = Z_{0,n}(0,t)$, then the system $\{X_n\}_{n=0}^{\infty}$ solves

$$\begin{aligned} dX_n &= \left(X_{n-1} + \frac{1}{2}X_n\right)dt + X_n dB_n \\ X_n(0) &= 1_{\{n=0\}}. \end{aligned}$$

in the Ito sense. In particular, if we define $Y_n(t) = e^{-\frac{3}{2}t}X_n(t)$ then $Y_n(t)$ solves

$$\begin{aligned} dY_n &= (Y_{n-1} - Y_n)\,dt + Y_n dB_n = -\nabla Y_n dt + Y_n dB_n \\ Y_n(0) &= 1_{\{n=0\}}. \end{aligned}$$

where ∇ is the forward difference operator on $\mathbb{Z}_+$. The last expression shows that Y_n can be viewed as the solution of the following 'totally asymmetric' analogue of the stochastic heat equation:

$$\begin{aligned} \partial_t Y &= -\nabla Y + YW \\ Y(0,n) &= \delta_{\{n=0\}} \end{aligned} \tag{2.2.4}$$

where W is space-time white noise on $\mathbb{R}_+ \times \mathbb{Z}_+$. Up to a deterministic multiplicative factor, we may therefore view the partition function (2.2.2) as giving the Feynman-Kac solution to the totally asymmetric stochastic heat equation (2.2.4), where the discrete Laplacian has been replaced by the forward difference operator. In [16], the authors studied the partition function in this model by taking this perspective. Using an analogue of the coordinate Bethe ansatz, they computed a contour integral representation for the integer moments of $Z_n(\beta)$. Asymptotic analysis then allowed them to compute the integer moment Lyapunov exponents.

Theorem ([16], Theorem 1.8)**.** *For any $s > 0$ and $k \in \mathbb{N}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log E\left[Z_{0,\lfloor ns \rfloor}(0,n)^k\right] = \min_{z>0} \left\{ \frac{k^2}{2} + kz - s \log \frac{\Gamma(z+k)}{\Gamma(z)} \right\}.$$

Note that the result in [16] corresponds to a limit of $Y_{\lfloor n\nu \rfloor}(n)$, which accounts for the extra factor of $-\frac{3}{2}k$ in the statement of the theorem in that paper. In [16, Appendix A], the authors conjectured that this result should extend to $k > 0$ as part of a verification that the replica computation of the free energy recovers the rigorous result of [69] for this model. Recalling the notation $\rho(s,t)$ from Lemma (2.2.1), the main result of the author's paper [49] is that this conjecture is correct. The following results are [49, Theorems 2.2 and 2.3].

Theorem 2.2.2. *For any $s, t > 0$ and $\lambda \in \mathbb{R}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log E\left[Z_{0,\lfloor ns \rfloor}(0,nt)^\lambda\right] := \Lambda_{s,t}(\lambda) = \begin{cases} \lambda\rho(s,t) & \lambda \leqslant 0 \\ \min_{z>0} \left\{ t\left(\frac{\lambda^2}{2} + \lambda z\right) - s \log \frac{\Gamma(z+\lambda)}{\Gamma(z)} \right\} & \lambda > 0 \end{cases}.$$

An application of the Gärtner-Ellis theorem then leads to the following result.

Theorem 2.2.3. *For any $s, t > 0$, the distributions of $n^{-1} \log Z_{0,\lfloor ns \rfloor}(0,nt)$ satisfy a large deviation principle with rate n and convex good rate function*

$$I_{s,t}(r) = \begin{cases} \infty & r < \rho(s,t) \\ \max_{\lambda,z>0} \left\{ r\lambda - t\left(\frac{\lambda^2}{2} + \lambda z\right) + s \log \frac{\Gamma(z+\lambda)}{\Gamma(z)} \right\} & r \geqslant \rho(s,t) \end{cases}.$$

One can check that the max and min in the previous expressions have unique extremizers by checking that the functions in question are strictly convex and concave respectively and have compact sub- and super-level sets.

Remark 2.2.4. Take $r > \rho(s,t) = \min\{t\theta - s\Psi_0(\theta)\}$, so that

$$I_{s,t}(r) = \max_{\lambda, z > 0} \left\{ r\lambda - t\left(\frac{\lambda^2}{2} + \lambda z\right) + s \log \frac{\Gamma(z+\lambda)}{\Gamma(z)} \right\}$$

The minimizing pair $(\mathrm{z}_\star, \lambda_\star) := (\mathrm{z}_\star(r), \lambda_\star(r))$ solves

$$r = t(\lambda_\star + \mathrm{z}_\star) - s\Psi_0(\mathrm{z}_\star + \lambda_\star), \qquad 0 = -t\,\lambda_\star + s\Psi_0(\mathrm{z}_\star + \lambda_\star) - s\Psi_0(\mathrm{z}_\star).$$

Combining these conditions, we see that $\mathrm{z}_\star$ and $\mathrm{z}_\star + \lambda_\star$ both solve

$$r = t(\mathrm{z}_\star + \lambda_\star) - s\Psi_0(\mathrm{z}_\star + \lambda_\star), \qquad r = t\,\mathrm{z}_\star - s\Psi_0(\mathrm{z}_\star).$$

This system has an interpretation: $t\theta - s\Psi_0(\theta)$ is the free energy in the stationary point to point O'Connell-Yor polymer with parameter θ, which will be introduced shortly. The function $\theta \mapsto t\theta - s\Psi_0(\theta)$ is strictly convex with a unique minimum at $\theta = \Psi_1^{-1}(t/s)$ and, as noted above, at this point it is equal to the shape function $\rho(s,t)$. To find the minimizers of the rate function, one then finds the two solutions to $r = tz - s\Psi_0(z)$. The smaller of the two solutions is $\mathrm{z}_\star$ and the difference between the solutions is $\lambda_\star$. Because $tz - s\Psi_0(z)$ is minimized at $\mathrm{z}_\star(0) := \Psi_1^{-1}(t/s)$ with value $\rho(s,t)$, we have

$$t(\mathrm{z}_\star(0) + \delta) - s\Psi_0(\mathrm{z}_\star(0) + \delta) = \rho(s,t) - \frac{1}{2}s\Psi_2(\mathrm{z}_\star(0))\delta^2 + o(\delta^2).$$

From this we see that for $\epsilon > 0$ small, the solutions to $\rho(s,t) + \epsilon = tz - s\Psi_0(z)$ are given by

$$\mathrm{z}_\star(0) \pm \sqrt{-\frac{1}{2}s\Psi_2(\mathrm{z}_\star(0))}^{-1}\sqrt{\epsilon} + o(\sqrt{\epsilon}).$$

It follows that

$$\lambda_\star(\rho(s,t) + \epsilon) = 2\sqrt{-\frac{1}{2}s\Psi_2(\mathrm{z}_\star(0))}^{-1}\sqrt{\epsilon} + o(\sqrt{\epsilon}).$$

By convex duality, we also have a representation $I_{s,t}(\rho(s,t)+\epsilon) = \int_0^\epsilon \lambda_\star(\rho(s,t)+x)dx$, from which it follows that

$$\begin{aligned} I_{s,t}(\rho(s,t)+\epsilon) &= \int_0^\epsilon \lambda_\star(\rho(s,t)+x)dx \\ &= \frac{4}{3}\sqrt{-\frac{1}{2}s\Psi_2(\Psi_1^{-1}(t/s))}^{-1}\epsilon^{\frac{3}{2}} + o\left(\epsilon^{\frac{3}{2}}\right). \end{aligned}$$

This is formally consistent with the observed $n^{1/3}$ fluctuations and the leading order right tail asymptotics of the Tracy-Widom GUE distribution. As noted above, this limit was proven in [17, Theorem 1.3].

The proofs of Theorems 2.2.2 and 2.2.3 follow an approach introduced by Seppäläinen in [81] and subsequently applied in [80, 37, 32]. Georgiou and Seppäläinen used this method to compute the large deviation rate function with normalization n for the free energy in the related log-gamma polymer in [37]. The key technical condition making this scheme tractable is the independence provided by the Burke property, which the log-gamma polymer shares with the O'Connell-Yor polymer.

The Burke property and the stationary O'Connell-Yor model

For $\theta > 0$, $t \in \mathbb{R}$ and $n \in \mathbb{Z}_+$ define point-to-point partition functions by

$$Z_n^\theta(t) = \int\limits_{-\infty<u_0<u_1<\cdots<u_{n-1}<t} e^{\theta u_0 - B_0(u_0) + B_1(u_0,u_1)+\cdots+B_n(u_{n-1},t)}du_0\ldots du_{n-1},$$

with the convention that

$$Z_0^\theta(t) = e^{\theta t - B_0(t)}.$$

We can think of $Z_n^\theta(t)$ as a modification of the polymer in the previous subsection where we add a spatial dimension, start in the infinite past, and modify the Brownian potential on line zero. Sample paths η in this modified model are non-decreasing, take values in $\mathbb{Z}_+$, have jumps of size one, and satisfy $\eta(s) = 0$ for all s sufficiently small, $\eta(t) = n$. See Figure 5.

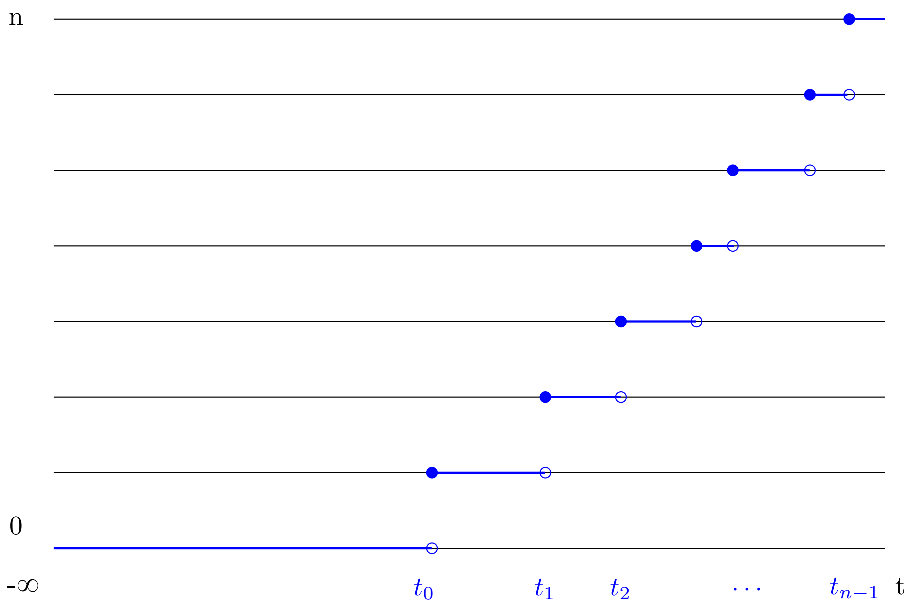


Figure 5: A sample path in the stationary O'Connell-Yor polymer model

For $s, t > 0$ and n sufficiently large that $ns \geqslant 1$, we obtain a decomposition of $Z^{\theta}_{[ns]}(nt)$ into terms that involve the partition functions we are studying by considering where paths leave the potential of the Brownian motion B:

$$Z^{\theta}_{[ns]}(nt) = \int_0^{nt} Z^{\theta}_0(u) Z_{1,[ns]}(u, nt) du + \sum_{j=1}^{[ns]} Z^{\theta}_j(0) Z_{j,[ns]}(0, nt). \tag{2.2.5}$$

This expression also leads to the interpretation of $Z^{\theta}_n(t)$ as a modification of the point-to-point partition function discussed in the previous subsection where we have added boundary conditions.

We will refer to this model as the stationary polymer, where the term stationary comes from the fact that it satisfies an analogue of Burke's theorem for M/M/1 queues. This fact is one of the main contributions of [71] and we refer the reader to that paper for a more in depth discussion of the connections to queueing theory. We follow the notation of [83], which contains the version of the Burke property that is used in this paper. Define $Y^{\theta}_0(t) = B(t)$ and for $k \geqslant 1$ recursively set

$$\begin{aligned}
r_k^\theta(t) &= \log \int_{-\infty}^t e^{Y_{k-1}^\theta(u,t)-\theta(t-u)+B_k(u,t)} du, \\
Y_k^\theta(t) &= Y_{k-1}^\theta(t) + r_k^\theta(0) - r_k^\theta(t), \qquad (2.2.6)\\
X_k^\theta(t) &= B_k(t) + r_k^\theta(0) - r_k^\theta(t);
\end{aligned}$$

then we have

Lemma 2.2.5 ([83], Theorem 3.3)**.** *Let $n \in \mathbb{N}$ and $0 \leqslant s_n \leqslant s_{n-1} \leqslant \cdots \leqslant s_1 < \infty$. Then over j, the following random variables and processes are all mutually independent.*

$$r_j(s_j) \quad \textit{and} \quad \{X_j(s) : s \leqslant s_j\} \quad \textit{for} \quad 1 \leqslant j \leqslant n, \quad \{Y_n(s) : s \leqslant s_n\},$$
$$\textit{and} \quad \{Y_j(s_{j+1}, s) : s_{j+1} \leqslant s \leqslant s_j\} \quad \textit{for } 1 \leqslant j \leqslant n-1.$$

Furthermore, the X_j and Y_j processes are standard Brownian motions, and $e^{-r_j(s_j)}$ is $\Gamma(\theta, 1)$ distributed.

An induction argument shows that

$$\sum_{k=1}^n r_k^\theta(t) = B(t) - \theta t + \log Z_n^\theta(t). \qquad (2.2.7)$$

As we will see shortly, expression (2.2.5) would lead to a variational formula for the right tail rate function we are looking for in terms of the right tail rate function of $Z_{\lfloor ns \rfloor}^\theta(nt)$. This right tail rate function would be tractable using (2.2.7) if $B(nt)$ were independent of $\sum_{k=1}^{\lfloor ns \rfloor} r_k^\theta(nt)$; as this is not the case, it is convenient to rewrite (2.2.5) in a form that separates these two terms:

$$e^{\sum_{k=1}^{\lfloor ns \rfloor} r_k^\theta(nt)} = n \int_0^t \frac{Z_0^\theta(nu)}{Z_0^\theta(nt)} Z_{1,\lfloor ns \rfloor}(nu, nt) du + \sum_{j=1}^{\lfloor ns \rfloor} \frac{Z_j^\theta(0)}{Z_0^\theta(nt)} Z_{j,\lfloor ns \rfloor}(0, nt). \qquad (2.2.8)$$

Having proven Theorem 2.2.2 using (2.2.8), we can take advantage of (2.2.5) to prove the corresponding result for the stationary model. This result can be compared with Theorem

2.2.11 and [37, Theorem 2.11], where the corresponding result for the stationary log gamma polymer was proven. The structure of the terms appearing in the maximum below is the same as in those results.

Theorem 2.2.6. *Fix* $\theta \in (0,\infty)$*, then for any* $\lambda \in (0,\infty)$

$$\Lambda^{\theta}_{s,t}(\lambda) := \lim_{n\to\infty} n^{-1}\log E\left[\left(Z^{\theta}_{\lfloor ns\rfloor}(nt)\right)^{\lambda}\right]$$
$$= \begin{cases} \left\{t\left(\frac{\lambda^2}{2}+\theta\lambda\right) - s\log\frac{\Gamma(\lambda+\theta)}{\Gamma(\theta)}\right\} \vee \left\{t\left(-\frac{\lambda^2}{2}+\theta\lambda\right) - s\log\frac{\Gamma(\theta-\lambda)}{\Gamma(\theta)}\right\} & \lambda < \theta \\ \infty & \lambda \geqslant \theta \end{cases}$$

Remark 2.2.7. For fixed s,t, the function

$$z \mapsto t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{\Gamma(z+\lambda)}{\Gamma(z)}$$

is strictly convex on $z \in (0,\infty)$ and has compact sublevel sets, so unique minimizers exist. The terms appearing in the maximum in Theorem 2.2.6 are the values of this function at $z = \theta$ and $z = \theta - \lambda$. It then follows that for $\Lambda_{s,t}$ as in Theorem 2.2.2 and $\Lambda^{\theta}_{s,t}$ as in Theorem 2.2.6 and fixed $\lambda > 0$

$$\min_{\theta>0} \Lambda^{\theta}_{s,t}(\lambda) > \Lambda_{s,t}(\lambda).$$

This is in constrast to the behavior of the free energies, where we have

$$\rho(s,t) = \min_{\theta>0}\{\theta t - s\Psi_0(\theta)\}$$

and where, by (2.2.7), $\theta t - s\Psi_0(\theta) = \lim n^{-1}\log Z^{\theta}_{\lfloor ns\rfloor}(nt)$. The same phenomenon is observed in the log gamma polymer [37, Remark 2.15] and for Brownian directed percolation in Remark 2.2.12.

2.2.2 Brownian directed percolation

Let $Z_n(\beta)$ be given by (2.2.1). One can see using Laplace's method that

$$\lim_{\beta\to\infty} \beta^{-1}\log Z_n(\beta) = \max_{0<s_1<\cdots<s_{n-1}<n}\{B_0(0,s_1) + \cdots + B_n(s_{n-1},n)\}.$$

The expression on the right hand side is the last passage time for a directed last-passage percolation model in a white noise random environment on $\mathbb{R}_+ \times \mathbb{Z}_+$, which we will call Brownian directed percolation. Here the paths are the same as paths in the O'Connell-Yor polymer; recall Figure 4. Introduce the notation $L_n(t)$ for this random variable:

$$L_n(t) = \max_{0<s_1<\cdots<s_{n-1}<t} \{B_0(0,s_1) + \cdots + B_n(s_n,t)\}\,. \tag{2.2.9}$$

As with the O'Connell-Yor polymer, it is convenient to have a family of last passage times for all point-to-point paths. To that effect, we define the last passage time from (u,k) to (t,n) by

$$L_{k,n}(u,t) = \sup_{u=s_{k-1}<s_k<\cdots<s_{n-1}<s_n=t} \left\{ \sum_{j=k}^{n} B_j(s_{j-1},s_j) \right\}.$$

A distributional equivalence between the last passage time $L_n(1)$ and the largest eigenvalue of a GUE matrix was discovered independently by Baryshnikov [7, Theorem 0.7] and Gravner, Tracy, and Widom [38], both in 2001. Although we will not use this fact, it is interesting to note that this extends to the process level. It is shown in [72] that $L_n(\cdot)$ has the same law as the largest eigenvalue process of an Hermitian Brownian motion. With this connection, the analogue of the free energy limit (in distribution and hence in probability) for this model follows from classical results in random matrix theory. The almost sure version of this limit for $L_n(\cdot)$ is due to Hambly, Martin, and O'Connell [44].

Theorem ([44], Theorem 8)**.** *Almost surely, for all $t > 0$,*

$$\lim_{n\to\infty} \frac{L_n(nt)}{n} = 2\sqrt{t}.$$

Note that by Brownian scaling, $L_n(t) \stackrel{d}{=} \sqrt{t}L_n(1)$. Using the distributional equivalence from [7, 38] and this scaling relation, the fluctuations around this limit correspond to the original Tracy-Widom GUE limit studied by Tracy and Widom in their seminal paper [87].

Theorem. *[87] For $t > 0$ and $r \in \mathbb{R}$,*

$$\lim_{n\to\infty} P\left(\frac{L_n(nt) - 2n\sqrt{t}}{\sqrt{t}n^{\frac{1}{3}}} \leqslant r\right) = F_2(r)$$

where $F_2(r)$ is the CDF of the Tracy-Widom GUE distribution.

Once again through the GUE connection, large deviation results are known at both rate n and n^2. These results depend on the large deviation principle for the empirical distribution of a Gaussian Unitary Ensemble matrix, which is due to Ben Arous and Guionnet [10, Theorem 1.3]. The right tail large deviation rate function can then be derived as in the computation of the corresponding rate function for the largest eigenvalue of a Gaussian Orthogonal Ensemble matrix in [9, Theorem 6.2]. The precise expression here is taken from the lecture notes of Ledoux on concentration inequalities for largest eigenvalues [64, (1.25)] and again uses Brownian scaling.

Theorem 2.2.8. *For any $r \geqslant 0$,*

$$\lim_{n\to\infty} -n^{-1} \log P\left(n^{-1}L_n(n) \geqslant 2 + r\right) = 4\int_0^{\frac{r}{2}} \sqrt{x(x+2)}dx$$

In Section 2.3, we present a fairly short proof of Theorem 2.2.8 using ideas which have previously been used to derive large deviation principles for the free energy of certain solvable positive and zero temperature directed polymer models in [32, 37, 49, 80, 81]. This approach avoids the GUE connection entirely and so provides a directed polymer proof of a result about a directed polymer model. This result plays a role in the computation of the rate n large deviation rate function for the O'Connell-Yor polymer free energy in [49] (and thus in Section 2.4). The argument in this paper then has the benefit of making the directed polymer large deviation literature a bit more self contained. The proof presented below shows that the limit in the statement of the theorem exists by subadditivity argument, from which we immediately derive the following corollary.

Corollary 2.2.9. *For any n and $r \geqslant 0$,*

$$P\left(n^{-1}L_n(n) \geqslant 2(1+r)\right) \leqslant e^{-nJ_{GUE}(r)}.$$

The result in Corollary 2.2.9 is also known [64, (2.6)] and can be derived via a weak limit procedure from the corresponding right tail estimates in [51, 80] (which can be derived as in this paper) for the i.i.d. exponential or geometric last passage percolation models. See the discussion after the statement of [64, Proposition 2.1]. Our approach is more direct. As in [51, 80], the result arises for free from the proof of Theorem 2.2.8. This implies a small deviation estimate [64, (2.7)] for the largest eigenvalue of a Gaussian Unitary Ensemble matrix of the type studied in [5, 65].

The key point making this polymer point of view tractable is the existence of an analogue of Burke's theorem from queueing theory for this model. This connection also implies the existence of a stationary polymer model, for which a result similar to Theorem 2.2.8 can be derived.

The Burke property for a stationary Brownian queue

As was the case for the O'Connell-Yor polymer, the key point which will make computation of the large deviation rate function tractable is an analogue of Burke's theorem from queueing theory. In fact, the result for Brownian directed percolation is a consequence of the classical Burke theorem after an application on Donsker's principle. See [45] or [71, Theorem 2]. Following the notation in [71], for each $\mu > 0$, we define

$$q_1(t) = \sup_{-\infty < s \leqslant t} \{B_0(s,t) + B_1(s,t) - \mu(t-s)\}$$

$$d_1(s,t) = B_0(s,t) + q_1(s) - q_1(t)$$

and recursively for $k \geqslant 2$

$$q_k^{\mu}(t) = \sup_{-\infty < s \leqslant t} \{d_{k-1}(s,t) + B_k(s,t) - \mu(t-s)\}$$

$$d_k(s,t) = d_{k-1}(s,t) + q_k^\mu(s) - q_k^\mu(t).$$

These have the interpretation as the 'departures' and 'queue length' processes for the server at station k in a stationary queueing Brownian queueing model. The version of Burke's theorem which we will need follows from a result in [71].

Theorem ([71], Theorem 2)**.** *For each $t \geqslant 0$, the family $\{q_k^\mu(t)\}_{k=1}^\infty$ consists of i.i.d. exponential random variables with mean μ^{-1}.*

We can extend the last pasage percolation model to paths like those considered in the stationary O'Connell-Yor model; see Figure 5. In this new environment, we define a family of last passage times by

$$\begin{aligned} L_n^\mu(t) &= \sup_{-\infty < s_0 < s_1 < \cdots < s_{n-1} < s_n = t} \left\{ \mu s_0 - B_0(s_0) + \sum_{j=1}^n B_j(s_{j-1}, s_j) \right\} \\ &= \sup_{-\infty < s_0 < t} \left\{ \mu s_0 - B_0(s_0) + L_{1,n}(s_0, t) \right\}. \end{aligned}$$

With these definitions, an induction argument shows that

$$\sum_{k=1}^n q_k^\mu(t) = B_0(t) - \mu t + L_n^\mu(t). \tag{2.2.10}$$

In particular, $\sum_{k=1}^n q_k^\mu(0) = L_n^\mu(0)$. We think of paths in this extended directed percolation model as being indexed by the points where they exit the lines $\{0, \ldots, n\}$. By grouping paths into those that exit line 0 before time 0 and those that exit after, we obtain

$$L_n^\mu(t) = \max_{0 \leqslant s_0 \leqslant t} \{\mu s_0 - B_0(s_0) + L_{1,n}(s_0, t)\} \vee \max_{1 \leqslant j \leqslant n} \left\{ L_j^\mu(0) + L_{j,n}(0,t) \right\}. \tag{2.2.11}$$

The decomposition in (2.2.11) can be viewed as describing a 'stationary' point-to-point polymer on $\mathbb{R}_+ \times \mathbb{Z}_+$ with i.i.d. exponential boundary conditions $\{L_{n+1}^\mu(0) - L_n^\mu(0)\}_{n \in \mathbb{N}}$ on the vertical axis and drifted Brownian boundary conditions $\{\mu t - B_0(t)\}_{t \geqslant 0}$ on the horizontal axis. Stationarity here is in the sense that, for example, $\{L_{n+1}^\mu(t) - L_n^\mu(t)\}_{n \in \mathbb{N}}$ is an i.i.d. exponential family for each $t > 0$. We will combine the queueing picture with this finite n variational problem in order to obtain a variational problem for the Lyapunov exponents in this model which will allow us to prove Theorem 2.2.8.

Remark 2.2.10. Take $r > 2$. We will show below that

$$\lim_{n\to\infty} -n^{-1}\log P\left(n^{-1}L_{\lfloor ns\rfloor}(nt) \geqslant r\right) = \max_{\lambda,z>0}\left\{\lambda r - t\left(\frac{1}{2}\lambda^2 + z\lambda\right) - s\log\frac{z+\lambda}{z}\right\}.$$

One can check that a unique minimizing pair $(\mathrm{z}_\star, \lambda_\star) := (\mathrm{z}_\star(r), \lambda_\star(r))$ exists and that this pair solves

$$r = t(\lambda_\star + \mathrm{z}_\star) + \frac{s}{\mathrm{z}_\star + \lambda_\star}, \qquad 0 = -t\,\lambda_\star - \frac{s}{\mathrm{z}_\star + \lambda_\star} + \frac{s}{\mathrm{z}_\star}.$$

We may combine these expressions to see that $\mathrm{z}_\star$ and $\mathrm{z}_\star + \lambda_\star$ solve

$$r = t(\lambda_\star + \mathrm{z}_\star) + \frac{s}{\mathrm{z}_\star + \lambda_\star}, \qquad r = t\,\mathrm{z}_\star + \frac{s}{\mathrm{z}_\star}.$$

This structure is the same as was observed in Remark 2.2.4 for the O'Connell Yor polymer. For fixed $\mu > 0$, $t\mu + s\mu^{-1}$ is the time constant in direction (s,t) for the stationary Brownian directed percolation model with parameter μ. To find the minimizers for the variational problem, one then finds the two solutions to $r = tz + sz^{-1}$. The smaller of the two is $\mathrm{z}_\star$ and the difference is $\lambda_\star$. We can compute these exactly and they are given by $\mathrm{z}_\star = (2t)^{-1}(r - \sqrt{r^2 - 4st})$ and $\lambda_\star = t^{-1}\sqrt{r^2 - 4st}$. Convex duality then implies that

$$\lim_{n\to\infty} -n^{-1}\log P\left(n^{-1}L_{\lfloor ns\rfloor}(nt) \geqslant r\right) = \int_{2\sqrt{st}}^{r} t^{-1}\sqrt{x^2 - 4st}\,dx.$$

Changing variables gives the expression in the statement of Theorem 2.2.8.

It is convenient to write (2.2.11) in a way that separates the terms $\sum_{k=1}^n q_k^\mu(t)$ and $B_0(t) - \mu t$, which are not independent:

$$\sum_{k=1}^{n} q_k^\mu(t) = \max_{0\leqslant s\leqslant t}\left\{\mu(s-t) + B_0(t) - B_0(s) + L_{1,n}(s,t)\right\} \tag{2.2.12}$$

$$\vee \max_{1\leqslant j\leqslant n}\left\{B_0(t)-\mu t+\sum_{k=1}^{j} q_k^{\mu}(0)+L_{j,n}(0,t)\right\}.$$

The key point in this decomposition is that for each $s_0 > 0$, the random variables $B_0(t)-B_0(s_0)$ and $L_{1,n}(s_0,t)$ are independent and for each $j \in \{1,\ldots,n\}$, the random variables $B_0(t)$, $\sum_{k=1}^{j} q_k^{\mu}(0)$, and $L_{j,n}(0,t)$ are mutually independent. This independence can be seen by recalling that the Brownian motions $\{B_i\}_{i=0}^{\infty}$ are independent and observing that $\sigma(B_i(s) : s \leqslant 0, i \in \mathbb{Z}_+)$ and $\sigma(B_i(s) : s \geqslant 0, i \in \mathbb{Z}_+)$ are independent. This decomposition will lead to a variational problem which can be used to prove Theorem 2.2.8. Once we have (re-)proven Theorem 2.2.8, we can bootstrap that result and the decomposition in (2.2.11) to compute the corresponding positive moment Lyapunov exponents for the stationary model. We will prove the following

Theorem 2.2.11. *For each $\mu, s, t > 0$ and $\lambda \geqslant 0$,*

$$\lim_{n\to\infty}\frac{1}{n}\log \mathrm{E}\left[e^{\lambda L^{\mu}_{\lfloor ns\rfloor}(nt)}\right] = \begin{cases}\left\{t\left(\frac{\lambda^2}{2}+\mu\lambda\right)+s\log\frac{\mu+\lambda}{\mu}\right\}\vee\left\{t\left(-\frac{\lambda^2}{2}+\mu\lambda\right)+s\log\frac{\mu}{\mu-\lambda}\right\} & \lambda<\mu \\ \infty & \lambda\geqslant\mu\end{cases}.$$

This result should be compared to Theorem 2.2.6 and [37, Theorem 2.11], where the corresponding result for the stationary log gamma polymer was proven. The structure of the terms appearing in the maximum above is the same as in those results.

Remark 2.2.12. Theorem 2.2.8 implies (and we show below) that for $s, t, \lambda > 0$

$$\lim_{n\to\infty}\frac{1}{n}\log \mathrm{E}\left[e^{\lambda L_{\lfloor ns\rfloor}(nt)}\right] = \min_{z>0}\left\{t\left(\frac{1}{2}\lambda^2+z\lambda\right)+s\log\frac{z+\lambda}{z}\right\},$$

where the function being minimized is strictly convex with a unique minimizer. The term appearing in Theorem 2.2.11 for $\lambda < \mu$ is the maximum of the values of this function at $z = \mu$ and $z = \mu - \lambda$. It follows that

$$\lim_{n\to\infty}\frac{1}{n}\log \mathrm{E}\left[e^{\lambda L_{\lfloor ns\rfloor}(nt)}\right] < \lim_{n\to\infty}\frac{1}{n}\log E\left[e^{\lambda L^{\mu}_{\lfloor ns\rfloor}(nt)}\right]$$

for all $\mu > \lambda$. This is in contrast to the behavior of the time constants, which satisfy (as is expected to be the case in general)

$$2\sqrt{st} = \min_{\mu>0} \left\{ \mu t + \frac{s}{\mu} \right\}.$$

The left hand side is the limit of $n^{-1}L_{\lfloor ns \rfloor}(nt)$ and the right hand side is the minimum of the limits of $n^{-1}L^{\mu}_{\lfloor ns \rfloor}(nt)$, where the minimizer is $\mu = \sqrt{s/t}$. The same phenomenon is observed in [37, Remark 2.15]. See also Remark 2.2.7.

Remark 2.2.13. By homogeneity, we may set $s = 1$. The condition for the right hand side of the expression in Theorem 2.2.11 to be the maximum of the two terms for a given $\lambda < \mu$ is $t\lambda^2 < -\log(1 - (\lambda/\mu)^2)$. Noting that $x < -\log(1 - x)$ for $0 < x < 1$, this is true for any such λ if $\mu \leqslant \sqrt{1/t}$. If $\mu > \sqrt{1/t}$, then there is a transition at the value of λ for which $t\lambda^2 = -\log(1-(\lambda/\mu)^2)$. This value can be expressed explicitly in terms of the principal branch of the Lambert W function.

An argument parallel to the proof of Theorem 2.2.11 also allows a computation of the corresponding right tail rate function. Denote the infimal convolution $f \,\square\, g(x) = \inf_y\{f(x - y) + g(y)\}$. We have

Theorem 2.2.14. *For all $s, t, \mu > 0$ and $x \in \mathbb{R}$, then*

$$J^{\mu}_{s,t}(x) := \lim_{n\to\infty} -n^{-1} \log P\left(L^{\mu}_{\lfloor ns \rfloor}(nt) \geqslant nx\right) = \inf_{0<r<t} \{g^{\mu}_r \,\square\, J_{s,t-r}(x)\} \wedge \inf_{0<u<s} \{h^{\mu}_u \,\square\, J_{s-u,t}(x)\}.$$

In particular, for $\mu \leqslant \sqrt{s/t}$,

$$J^{\mu}_{s,t}(x) := (2t)^{-1} \int_{t\mu+s\mu^{-1}}^{x} \sqrt{y^2 - 4st} + (2t\mu - y)dy 1_{\{x \geqslant t\mu+s\mu^{-1}\}}.$$

Remark 2.2.15. For simplicity restrict to $s = t = 1$ and consider only $0 < \mu \leqslant 1$. Then for such μ we have

$$\lim_{n\to\infty} -n^{-1} \log P\left(L^{\mu}_n(n) \geqslant nr\right) = \frac{1}{2} \int_{\mu+\mu^{-1}}^{r} \sqrt{x^2 - 4} - (x - 2\mu)dx 1_{\{r \geqslant \mu+\mu^{-1}\}}.$$

Substituting $r = \mu + \mu^{-1} + \epsilon$ for $\epsilon > 0$ small, the leading order small ϵ asymptotics of this function are $\mu^2(2(1-\mu^2))^{-1}\epsilon^2$ if $\mu < 1$ and $2/3\epsilon^{3/2}$ if $\mu = 1$. This is consistent with Gaussian fluctuations away from the characteristic direction and KPZ type fluctuations in the characteristic direction for the stationary model. The restrictions to $s = t = 1$ are without loss of generality because of homogeneity and the observation that for $\mu, t, a > 0$, $aL_n^{\mu}(t) \stackrel{d}{=} L_n^{\mu/a}(a^2t)$.

2.2.3 Inhomogeneous exponential last passage percolation

The first directed polymer models for which Conjecture 2.1.1 were verified are the zero temperature point-to-point (last passage percolation) models in which the environment $\{W(i,j)\}$ are i.i.d. with exponential or geometric marginals. This result is due to Johansson [51] and the same paper addresses left and right tail large deviations for these models. Models equivalent to these were studied prior to [51]. For example, through a combinatorial map these models can be shown to be equivalent to the totally asymmetric simple exclusion process (TASEP) run in continuous or discrete time respectively [82]. We will discuss the mapping connecting these models shortly. Since the models with i.i.d. exponential or geometric weights are so well understood, it is natural to ask whether one can relax some of the assumptions on the weights while still preserving exact solvability. This turns out to be the case.

Take two sequences $(\mathbf{a}, \mathbf{b})$ with $\mathbf{a} = (a_i)_{i \geqslant 1}$, $\mathbf{b} = (b_j)_{j \geqslant 1}$, and $a_i, b_j > 0$. If the family $\{W(i,j)\}$ are independent with the marginal distributions of $W(i,j)$ being either exponentially distributed with mean $(a_i+b_j)^{-1}$ or geometrically distributed with mean $e^{a_i+b_j}$, then the model remains exactly solvable. In fact, as observed by Johansson in 2001 [52], the distribution function of the last passage time in the geometric model can still be expressed in terms of the Schur measure introduced by Okounkov in [74]. Borodin and Peché [18] noted that one can take limits from geometric variables to exponential variables to obtain a continuous version of the Schur measure which gives the cumulative distribution function of the inhomogeneous exponential model. The same object appeared previously in [34].

One could use the explicit formulas coming from the connection to the Schur measure to prove large deviation results for these models. We take a different perspective. As mentioned above, the i.i.d. exponential and geometric models are connected to a variety of classical models in probability. Among these models is the M/M/1 queue. It was observed in [6, Theorem 3.1] that a version of Burke's theorem for the M/M/1 queue implies the existence of a last passage percolation model in which the vertical and horizontal increments of the last passage times have strong independence properties. This independence structure carries over to the inhomogeneous setting we consider.

In order to have interesting limit theorems, we need to place some restrictions on the parameters a_i and b_j. To see this, note if the parameters tend to zero too quickly (for example, if $a_i = b_i = i^{-2}$), then the normalized last passage times will not even be tight, much less have almost sure limits. It is natural in this situation to draw parameter sequences randomly from appropriate ergodic distributions. We consider two cases. We refer to the model where we condition on the parameter sequences as *quenched*. In this case, the results that follow are true under fairly mild ergodicity assumptions on the joint distribution of $(\mathbf{a}, \mathbf{b})$. We call the model obtained by averaging over the distribution of $(\mathbf{a}, \mathbf{b})$ *annealed*. In this case, our results are only valid under the assumptions that $\mathbf{a}$ and $\mathbf{b}$ are independent i.i.d. sequences.

In the quenched model described above, the weights at different sites are independent but (typically) not identically distributed exponential random variables. In contrast, in the annealed model weights along rows and columns share parameters and so are not independent. See Figure 6. For example, one can check directly that the covariance of $W(i, j)$ and $W(i, j')$ for $j \neq j'$ in the annealed model is $\mathrm{Var}(\mathrm{E}[(a_1 + b_1)^{-1} | a_1])$. This long-range dependence has a large impact on the behavior of the model.

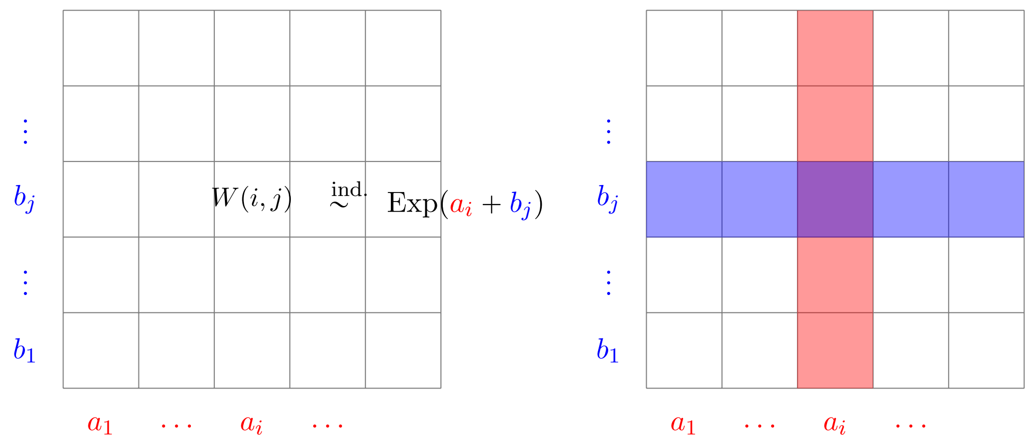


Figure 6: The inhomogeneous exponential environment, drawn as a growth model with weights assigned to lattice of squares, rather than the points of a lattice. In the annealed model, the weights share parameters along rows and columns and because of this are not independent.

In the language of particle systems, this corresponds to a totally asymmetric simple exclusion process with particle-wise and hole-wise inhomogeneity. The map which connects the two models is as follows. The exclusion process lives on $\mathbb{Z}$ and begins with particles at the sites $i \leqslant 0$ and holes at the sites $i > 0$. We label the particles from right to left and the holes from left to right with the natural numbers. The last passage time $G(m,n)$ is the time at which particle m and hole n interchange. In terms of dynamics, when particle m is immediately to the left of hole n, they interchange positions with rate $a_n + b_m$. One can check that the process defined in this way is Markov because the jump rates are exponential. The case when $a_i = b_i = \frac{1}{2}$ for all i is exactly the usual totally asymmetric simple exclusion process where particles move to the right at rate one, subject to the exclusion rule.

Point to point model

In the analysis of this model, we work in a cannonical setting. Denote by $W(i,j)$ the projection $\mathbb{R}_+^{\mathbb{N}^2} \to \mathbb{R}_+$ onto the coordinate (i,j) for $i,j \in \mathbb{N}$. For any sequences $\mathbf{a} = (a_1, a_2, \ldots), \mathbf{b} =$

$(b_1, b_2, \dots)$ taking values in $(0, \infty)$, we define $\mathbf{P}_{\mathbf{a},\mathbf{b}}$ to be the product measure on $\mathbb{R}_+^{\mathbb{N}^2}$ satisfying

$$\mathbf{P}_{\mathbf{a},\mathbf{b}}(W(i,j) \geqslant x) = e^{-(a_i+b_j)x} \quad \text{for } i,j \in \mathbb{N} \text{ and } x \geqslant 0.$$

We will draw the sequences $(\mathbf{a}, \mathbf{b})$ randomly from a distribution μ on $\mathbb{R}_+^{\mathbb{N}} \times \mathbb{R}_+^{\mathbb{N}}$. For $k \in \mathbb{Z}_+$, let τ_k denote the shift $(c_n)_{n\in\mathbb{N}} \mapsto (c_{n+k})_{n\in\mathbb{N}}$. In all of the results that follow, we make the following assumptions on $(\mathbf{a}, \mathbf{b})$. We assume that $\mathbf{a}$ and $\mathbf{b}$ are stationary sequences under μ. We assume further that μ is separately ergodic with respect to $\tau_k \times \tau_l$ for $k, l \in \mathbb{N}$. This means that if $k, l \in \mathbb{N}$ and $B \subset \mathbb{R}_+^{\mathbb{N}} \times \mathbb{R}_+^{\mathbb{N}}$ is a Borel set with $(\tau_k \times \tau_l)^{-1}(B) = B$ then $\mu(B) \in \{0, 1\}$.

The annealed distribution $\mathbb{P}$ is given by $\mathbb{P}(B) = \mathrm{E}\left[\mathbf{P}_{\mathbf{a},\mathbf{b}}(B)\right]$ for any Borel set $B \subset \mathbb{R}_+^{\mathbb{N}^2}$, where E is the expectation under μ. Let $\mathbf{E}_{\mathbf{a},\mathbf{b}}$ and $\mathbb{E}$ denote the expectations under $\mathbf{P}_{\mathbf{a},\mathbf{b}}$ and $\mathbb{P}$, respectively. We denote by α and β the distributions of a_1 and b_1 and take the convention that a and b are random variables with distributions α and β respectively. In all of the following results, we will assume that $\mathrm{E}[a + b] < \infty$ and $\underline{\alpha} + \underline{\beta} > 0$. Finally, all large deviation results under $\mathbb{P}$ are limited to the case where $\mathbf{a}$ and $\mathbf{b}$ are independent i.i.d. sequences. We denote the last passage time by

$$G(m,n) = \max_{\pi \in \Pi_{(1,1),(m,n)}} \sum_{(i,j)\in\pi} W(i,j). \tag{2.2.13}$$

where $\Pi_{(k,l),(m,n)}$ is the set all sequences $\pi = (u_i, v_i)_{i\in[p]}$ in $\mathbb{Z}^2$ such that $(u_1, v_1) = (k, l)$, $(u_p, v_p) = (m, n)$ and $(u_{i+1} - u_i, v_{i+1} - v_i) \in \{(1, 0), (0, 1)\}$ for $1 \leqslant i < p$. The change of notation in this section is because we use $\mathbf{L}$ and $\mathbb{L}$ to denote Lyapunov exponents.

We briefly summarize the results from [31]. The ergodicity assumptions on μ and the superadditivity of the last-passage times imply that $\lim_{n\to\infty} n^{-1} G(\lfloor ns \rfloor, \lfloor nt \rfloor) = g(s,t)$ for $s, t > 0$ $\mathbb{P}$-a.s. and $\mathbf{P}_{\mathbf{a},\mathbf{b}}$-a.s. for μ-a.e. $(\mathbf{a}, \mathbf{b})$ for some deterministic function g known as the *shape function.* g admits the variational representation

$$g(s,t) = \inf_{z\in[-\underline{\alpha},\underline{\beta}]} \left\{ s\,\mathrm{E}\left[\frac{1}{a+z}\right] + t\,\mathrm{E}\left[\frac{1}{b-z}\right] \right\} \quad \text{for } s,t > 0. \tag{2.2.14}$$

It is shown in [31] that the infimum above is actually a minimum with a unique minimizer and the function g_z given by $g_z(s,t) = s\,\mathrm{E}\left[(a+z)^{-1}\right] + t\,\mathrm{E}\left[(b-z)^{-1}\right]$ is the shape function of a stationary version of the model. At times we will also view $g(s,t)$ as a function of $(\alpha,\beta) \in \mathcal{M}_1(\mathbb{R}_+)^2$. In these cases, we will use the notation $(\alpha,\beta) \mapsto g(s,t) \equiv g_{\alpha,\beta}(s,t)$ to highlight the dependence on these measures. This map will be considered for any $(\alpha,\beta) \in \mathcal{M}_1(\mathbb{R}_+)^2$.

Set

$$c_1 = \frac{\mathrm{E}\left[(b+\underline{\alpha})^{-2}\right]}{\mathrm{E}\left[(a-\underline{\alpha})^{-2}\right]} \qquad c_2 = \frac{\mathrm{E}\left[(b-\underline{\beta})^{-2}\right]}{\mathrm{E}\left[(a+\underline{\beta})^{-2}\right]}. \tag{2.2.15}$$

Then $0 \leqslant c_1 < c_2 \leqslant \infty$, and $c_1 = 0$ and $c_2 = \infty$ if and only if $\mathrm{E}[(a-\underline{\alpha})^{-2}] = \infty$ and $\mathrm{E}[(b-\underline{\beta})^{-2}] = \infty$, respectively. It can be seen from (2.2.14) that g is strictly concave for $c_1 < s/t < c_2$ and is linear for $s/t \leqslant c_1$ or $s/t \geqslant c_2$, see Figure 7.

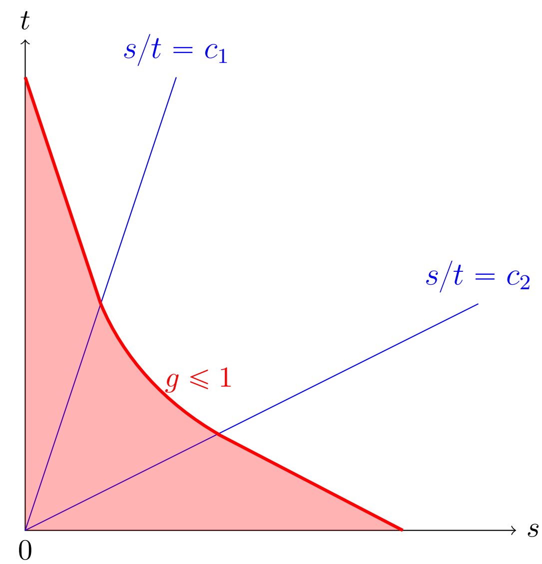


Figure 7: An illustration of the sublevel set $g \leqslant 1$ and the rays $s/t = c_1$ and $s/t = c_2$ when $0 < c_1 < c_2 < \infty$.

We show in Proposition 2.5.15 that for $s,t,\lambda > 0$, we may define the quenched and annealed Lyapunov exponents by

$$\mathbf{L}_{s,t}(\lambda) = \lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda G(\lfloor ns\rfloor,\lfloor nt\rfloor)}\right] \quad \mu\text{-a.s.,} \qquad \mathbb{L}_{s,t}(\lambda) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{\lambda G(\lfloor ns\rfloor,\lfloor nt\rfloor)}\right].$$

Our first result is an exact computation of these exponents.

Theorem 2.2.16. *For $s, t, \lambda > 0$,*

$$\mathbf{L}_{s,t}(\lambda) = \begin{cases} \inf\limits_{z \in [-\underline{\alpha}, \underline{\beta}-\lambda]} \left\{ s\,\mathrm{E}\log\left(\dfrac{a+z+\lambda}{a+z}\right) + t\,\mathrm{E}\log\left(\dfrac{b-z}{b-z-\lambda}\right)\right\} & \text{if } 0 < \lambda \leqslant \underline{\alpha} + \underline{\beta} \\ \infty & \text{if } \lambda > \underline{\alpha} + \underline{\beta}. \end{cases} \tag{2.2.16}$$

$$\mathbb{L}_{s,t}(\lambda) = \begin{cases} \inf\limits_{z \in [-\underline{\alpha}, \underline{\beta}-\lambda]} \left\{ s\log\mathrm{E}\left(\dfrac{a+z+\lambda}{a+z}\right) + t\log\mathrm{E}\left(\dfrac{b-z}{b-z-\lambda}\right)\right\} & \text{if } 0 < \lambda \leqslant \underline{\alpha} + \underline{\beta} \\ \infty & \text{if } \lambda > \underline{\alpha} + \underline{\beta} \end{cases} \tag{2.2.17}$$

Once we have Theorem 2.2.16, a proof similar to the proof of Theorem 2.2.16 allows us to compute the Lyapunov exponents in a stationary version of the model, which will be introduced in the next section. For the moment, we record the result and comment briefly on its implications.

Theorem 2.2.17. *For $z \in (-\underline{\alpha}, \underline{\beta})$, almost surely for all $s, t > 0$ and $\lambda \in (0, (\underline{\alpha}+z) \wedge (\underline{\beta}-z))$*

$$\begin{aligned} &\mathbf{L}^z_{s,t}(\lambda) := \lim_{n\to\infty} n^{-1}\log\mathbf{E}^z_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\hat{G}(\lfloor ns\rfloor, \lfloor nt\rfloor)}\right] \\ &= \left\{ s\,\mathrm{E}\left[\log\frac{a+z}{a+z-\lambda}\right] + t\,\mathrm{E}\left[\log\frac{b-z+\lambda}{b-z}\right]\right\} \vee \left\{ s\,\mathrm{E}\left[\log\frac{a+z+\lambda}{a+z}\right] + t\,\mathrm{E}\left[\log\frac{b-z}{b-z-\lambda}\right]\right\}. \end{aligned}$$

Similarly, we show in Proposition 2.5.14 that for $s, t > 0$ and $r \in \mathbb{R}$, we may define right tail rate functions by

$$\begin{aligned} &\lim_{n\to\infty} -\frac{1}{n}\log\mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor ns\rfloor, \lfloor nt\rfloor) \geqslant nr) = \mathbf{J}_{s,t}(r) \quad \mu\text{-a.s.}, \\ &\lim_{n\to\infty} -\frac{1}{n}\log\mathbb{P}(G(\lfloor ns\rfloor, \lfloor nt\rfloor) \geqslant nr) = \mathbb{J}_{s,t}(r) \end{aligned}$$

Using the previous result, we show that

Theorem 2.2.18. *For $s, t > 0$,*

$$\mathbf{J}_{s,t}(r) = \begin{cases} \sup\limits_{\substack{\lambda \in (0, \underline{\alpha}+\underline{\beta}] \\ z \in [-\underline{\alpha}, \underline{\beta}-\lambda]}} \left\{ r\lambda - s\,\mathrm{E}\log\left(\frac{a+z+\lambda}{a+z}\right) - t\,\mathrm{E}\log\left(\frac{b-z}{b-z-\lambda}\right)\right\} & r \geqslant g(s,t) \\ 0 & r < g(s,t) \end{cases} \tag{2.2.18}$$

$$\mathbb{J}_{s,t}(r) = \begin{cases} \sup\limits_{\substack{\lambda \in (0, \underline{\alpha}+\underline{\beta}] \\ z \in [-\underline{\alpha}, \underline{\beta}-\lambda]}} \left\{ r\lambda - s\log\mathrm{E}\left[\frac{a+z+\lambda}{a+z}\right] - t\log\mathrm{E}\left[\frac{b-z}{b-z-\lambda}\right]\right\} & r \geqslant g(s,t) \\ 0 & r < g(s,t) \end{cases} \tag{2.2.19}$$

As with the shape function, we will at times consider the maps $(\alpha, \beta) \mapsto \mathbf{J}_{s,t}(r) \equiv \mathbf{J}_{s,t}^{\alpha,\beta}(r)$ and $(\alpha, \beta) \mapsto \mathbb{J}_{s,t}(r) \equiv \mathbb{J}_{s,t}^{\alpha,\beta}(r)$.

Note that the Lyapunov exponents and the right tail rate functions depend on μ only through the marginal distributions α and β. The variational problem in (2.2.18) can be solved exactly for certain choices of α, β, s and t. We note that if $r \geqslant g(s,t)$ and there exists $\lambda_\star \in (0, \underline{\alpha}+\underline{\beta})$ and $\mathrm{z}_\star \in (-\underline{\alpha}, \underline{\beta}-\lambda_\star)$ such that

$$\begin{aligned} 0 &= s\,\mathrm{E}\left[\frac{1}{a+\mathrm{z}_\star+\lambda_\star} - \frac{1}{a+\mathrm{z}_\star}\right] + t\,\mathrm{E}\left[\frac{1}{b-\mathrm{z}_\star-\lambda_\star} - \frac{1}{b-\mathrm{z}_\star}\right] \\ r &= s\,\mathrm{E}\left[\frac{1}{a+\mathrm{z}_\star+\lambda_\star}\right] + t\,\mathrm{E}\left[\frac{1}{b-\mathrm{z}_\star-\lambda_\star}\right], \end{aligned}$$

then

$$\mathbf{J}_{s,t}(r) = \lambda_\star r - s\,\mathrm{E}\log\left(\frac{a+\mathrm{z}_\star+\lambda_\star}{a+\mathrm{z}_\star}\right) + t\,\mathrm{E}\log\left(\frac{b-\mathrm{z}_\star}{b-\mathrm{z}_\star-\lambda_\star}\right). \tag{2.2.20}$$

Example 2.2.19. If $\alpha = \beta = \delta_{c/2}$ for $c > 0$, then for $r \geqslant g(s,t) = c^{-1}(\sqrt{s}+\sqrt{t})^2$,

$$\mathbf{J}_{s,t}(r) = \sqrt{(s+t-cr)^2 - 4st} - 2s\cosh^{-1}\left(\frac{s-t+cr}{2\sqrt{csr}}\right) - 2t\cosh^{-1}\left(\frac{t-s+cr}{2\sqrt{ctr}}\right), \tag{2.2.21}$$

which recovers [80, Theorem 4.4].

Example 2.2.20. If $\alpha = \beta = p\delta_c + q\delta_d$ for $p, q, c, d > 0$ with $p + q = 1$ and $s = t$, then for $r \geqslant g(s,s) = 2s\left(pc^{-1} + qd^{-1}\right)$,

$$\begin{aligned}\mathbf{J}_{s,s}(r) = r\,\lambda_\star - &sp\log\left(\frac{c + \mathrm{z}_\star + \lambda_\star}{c + \mathrm{z}_\star}\right) - tq\log\left(\frac{c - \mathrm{z}_\star}{c - \mathrm{z}_\star - \lambda_\star}\right)\\ &- sq\log\left(\frac{d + \mathrm{z}_\star + \lambda_\star}{d + \mathrm{z}_\star}\right) - tq\log\left(\frac{d - \mathrm{z}_\star}{d - \mathrm{z}_\star - \lambda_\star}\right)\end{aligned}$$

where

$$\begin{aligned}\mathrm{z}_\star &= \frac{2cp + 2dq + c^2r + d^2r - \sqrt{\Delta}}{2r}, \qquad \mathrm{z}_\star + \lambda_\star = \frac{2cp + 2dq + c^2r + d^2r + \sqrt{\Delta}}{2r},\\ \Delta &= (2cp + 2dq + c^2r + d^2r)^2 + 4r(2cd^2p + 2c^2dq - c^2d^2r).\end{aligned}$$

More complicated exact formulas in this model are available in all directions (s,t).

Example 2.2.21. If α and β are uniform on $[c/2, c/2 + l]$ for $c, l > 0$ and $s = t$, then

$$\mathbf{J}_{s,s}(r) = r\,\lambda_\star - \frac{2s}{l}\int_{c/2}^{c/2+l}\log\left(\frac{x + \mathrm{z}_\star + \lambda_\star}{x + \mathrm{z}_\star}\right)dx \quad \text{for } r \geqslant g(s,s) = \frac{2s}{l}\log\left(1 + \frac{2l}{c}\right),$$

where

$$\mathrm{z}_\star = -\sqrt{\frac{(c/2 + l)^2 - c^2e^{rl/s}/4}{1 - e^{rl/s}}} \qquad \mathrm{z}_\star + \lambda_\star = \sqrt{\frac{(c/2 + l)^2 - c^2e^{rl/s}/4}{1 - e^{rl/s}}}.$$

Left tail large deviations in the quenched model have rate strictly larger than n. We expect that under mild hypotheses the correct rate should be n^2, as is the case in the homogeneous model where $\alpha = \beta = \delta_{\frac{c}{2}}$ [51, 80].

Lemma 2.2.22. $\lim\limits_{n\to\infty} -\dfrac{1}{n}\log\mathbf{P}_{\mathbf{a},\mathbf{b}}\left(G(\lfloor ns\rfloor, \lfloor nt\rfloor) \leqslant nr\right) = \infty$ *for* $s, t > 0$ *and* $r < g(s,t)$ μ*-a.s.*

Combining our results for the right and left tail deviations, we can prove a full quenched LDP at rate n. The rate function is given by

$$\mathbf{I}_{s,t}(r) = \begin{cases}\mathbf{J}_{s,t}(r) & r \geqslant g(s,t)\\ \\ \infty & r < g(s,t)\end{cases}. \tag{2.2.22}$$

As before, we will at times use the notation $(\alpha,\beta) \mapsto \mathbf{I}_{s,t}(r) \equiv \mathbf{I}_{s,t}^{\alpha,\beta}(r)$.

Theorem 2.2.23. *μ-a.s, for any $s, t > 0$, the distribution of $n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor)$ under $\mathbf{P}_{\mathbf{a},\mathbf{b}}$ satisfies a large deviation principle with rate n and convex, good rate function $\mathbf{I}_{s,t}$.*

Although our proof of the large deviation principle goes through the Lyapunov exponents, we do not apply the Gärtner-Ellis theorem. The steepness condition in this model is $\mathrm{E}[(a - \underline{\alpha})^{-1}] = \mathrm{E}[(b - \underline{\beta})^{-1}] = \infty$, which would rule out having linear segments of the shape function and so is too restrictive.

In contrast to the quenched case, there are non-trivial annealed large deviations at rate n. The following bound gives a mechanism for these deviations. In the statement $H(\cdot|\cdot)$ denotes the relative entropy.

Lemma 2.2.24. *For any $x < y$,*

$$\limsup_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor) \in (x,y)) \leqslant \inf_{\substack{\nu_1 \in \mathcal{M}^{\alpha}, \nu_2 \in \mathcal{M}^{\beta} \\ g_{\nu_1,\nu_2}(s,t) \in (x,y)}} \{s\,\mathrm{H}(\nu_1|\alpha) + t\,\mathrm{H}(\nu_2|\beta)\}$$

The other bound needed to show that n is the correct rate for certain left tail large deviations follows from essentially the same argument used to show that the quenched rate is strictly larger than n. This is discussed briefly after the proof of Lemma 2.2.22. To show that there are rate n annealed left tail large deviations it suffices to show that there exist $\nu_1 \in \mathcal{M}^{\alpha}$ and $\nu_2 \in \mathcal{M}^{\beta}$ with $g_{\nu_1,\nu_2}(s,t) < g_{\alpha,\beta}(s,t)$. We give a simple proof that under mild conditions this is the case in Lemma 2.5.11. We expect that this mechanism is not sharp.

Example 2.2.25. Suppose that $\alpha = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_2$ and $\beta = \delta_1$, and recall that $\mathcal{M}^{\alpha} = \{p\delta_1 + (1-p)\delta_2 : 0 \leqslant p \leqslant 1\}$. For $0 \leqslant p \leqslant 1$, call $\alpha_p = p\delta_1 + (1-p)\delta_2$. Then $\{g_{\alpha_p,\beta}(1,9) : 0 \leqslant p \leqslant 1\} = \{5.\bar{3}\} \cup (5.5, 8]$. The reason for the discontinuity in this example is that if $p > 0$, then the functional in (2.2.14) is minimized on the set $(-1, 1)$, but if $p = 0$, the minimization occurs on $(-2, 1)$. We have chosen $s = 1, t = 9$ so that the minimizer for the $p = 0$ case occurs in $(-2, -1)$. The bound one obtains from Lemma 2.2.24 in this example is infinite when applied to the interval $(5.4, 5.5)$. The finite relative entropy perturbation of the a_i parameters switching the distribution to δ_2 turns this into a right tail large deviation.

The next theorem connects quenched rate function and annealed right tail rate function through a variational problem. We expect that this result means that large deviations above the shape function in the annealed model with marginals α and β can be viewed as a large deviation in the parameters $\{a_i\}_{i=1}^{\lfloor ns \rfloor}$ and $\{b_j\}_{j=1}^{\lfloor nt \rfloor}$ which affect the distribution of $G(\lfloor ns \rfloor, \lfloor nt \rfloor)$, followed by a deviation in the quenched model with these perturbed parameters. Our proof is purely analytic and does not show this interpretation directly. A similar, but stronger, connection was shown for random walk in a random environment by Comets, Gantert and Zeitouni in [22].

Theorem 2.2.26. *For any $s, t > 0$ and $r > g(s,t)$,*

$$\mathbb{J}_{s,t}^{\alpha,\beta}(r) = \inf_{\substack{\nu_1 \in \mathcal{M}^\alpha \\ \nu_2 \in \mathcal{M}^\beta}} \left\{ \mathbf{I}_{s,t}^{\nu_1,\nu_2}(r) + s\,\mathrm{H}(\nu_1|\alpha) + t\,\mathrm{H}(\nu_2|\beta) \right\}.$$

A minimizing pair (ν_1, ν_2) exists and the equality

$$\mathbb{J}_{s,t}^{\alpha,\beta}(r) = \mathbf{I}_{s,t}^{\nu_1,\nu_2}(r) + s\,\mathrm{H}(\nu_1|\alpha) + t\,\mathrm{H}(\nu_2|\beta)$$

holds if and only if

$$\frac{d\nu_1}{d\alpha}(a) = \frac{a + \mathrm{z}_\star + \lambda_\star}{a + \mathrm{z}_\star}\,\mathrm{E}\left[\frac{a + \mathrm{z}_\star + \lambda_\star}{a + \mathrm{z}_\star}\right]^{-1}, \qquad \frac{d\nu_2}{d\beta}(b) = \frac{b - \mathrm{z}_\star}{b - \mathrm{z}_\star - \lambda_\star}\,\mathrm{E}\left[\frac{b - \mathrm{z}_\star}{b - \mathrm{z}_\star - \lambda_\star}\right]^{-1}$$

where $\mathrm{z}_\star$ and $\lambda_\star$ are the unique $\mathrm{z}_\star, \lambda_\star$ with $\lambda_\star \in [0, \underline{\alpha} + \underline{\beta}], \mathrm{z}_\star \in [-\underline{\alpha}, \underline{\beta} - \lambda_\star]$ satisfying

$$\mathbb{J}_{s,t}^{\alpha,\beta}(r) = r\,\lambda_\star - s\log \mathrm{E}^\alpha\left[\frac{a + \mathrm{z}_\star + \lambda_\star}{a + \mathrm{z}_\star}\right] - t\log \mathrm{E}^\beta\left[\frac{b - \mathrm{z}_\star}{b - \mathrm{z}_\star - \lambda_\star}\right]. \tag{2.2.23}$$

It is natural to conjecture that this variational connection describes all rate n annealed large deviations, rather than just annealed right tail large deviations. We have been unable to prove this result.

The next result concerns the regularity of our rate functions. Our rate functions are convex and differentiable to the right of $g(s,t)$, but we note that for certain choices of α and β they can have linear segments; see Lemma 2.5.9 and the comments preceding it.

Theorem 2.2.27. *For any $s,t>0$, both $\mathbf{J}_{s,t}$ and $\mathbb{J}_{s,t}$ are continuously differentiable on $[g(s,t),+\infty)$.*

Finally, we describe the leading order asymptotics of $\mathbf{J}_{s,t}(r)$ and $\mathbb{J}_{s,t}(r)$ as $r\downarrow g(s,t)$ and comment on the implications for the fluctuations of the last-passage times. Let ζ denote the unique minimizer of (2.2.14).

Theorem 2.2.28. *For any $s,t>0$, as $\epsilon\downarrow 0$,*

$$\mathbf{J}_{s,t}(g(s,t)+\epsilon)=\begin{cases}\left(-s\,\mathrm{E}\left[\dfrac{2}{(a-\underline{\alpha})^2}\right]+t\,\mathrm{E}\left[\dfrac{2}{(b+\underline{\alpha})^2}\right]\right)^{-1}\epsilon^2+o(\epsilon^2) & \text{if } s/t<c_1\\ \dfrac{2}{3}\left(s\,\mathrm{E}\left[\dfrac{1}{(a-\underline{\alpha})^3}\right]+t\,\mathrm{E}\left[\dfrac{1}{(b+\underline{\alpha})^3}\right]\right)^{-1/2}\epsilon^{3/2}+o(\epsilon^{3/2}) & \text{if } s/t=c_1\\ & \text{and } \mathrm{E}[(a-\underline{\alpha})^{-3}]<\infty\\ \dfrac{4}{3}\left(s\,\mathrm{E}\left[\dfrac{1}{(a+\zeta)^3}\right]+t\,\mathrm{E}\left[\dfrac{1}{(b-\zeta)^3}\right]\right)^{-1/2}\epsilon^{3/2}+o(\epsilon^{3/2}) & \text{if } c_1<s/t<c_2\\ \dfrac{2}{3}\left(s\,\mathrm{E}\left[\dfrac{1}{(a+\underline{\beta})^3}\right]+t\,\mathrm{E}\left[\dfrac{1}{(b-\underline{\beta})^3}\right]\right)^{-1/2}\epsilon^{3/2}+o(\epsilon^{3/2}) & \text{if } s/t=c_2\\ & \text{and } \mathrm{E}[(b-\underline{\beta})^{-3}]<\infty\\ \left(s\,\mathrm{E}\left[\dfrac{2}{(a+\underline{\beta})^2}\right]-t\,\mathrm{E}\left[\dfrac{2}{(b-\underline{\beta})^2}\right]\right)^{-1}\epsilon^2+o(\epsilon^2) & \text{if } s/t>c_2\end{cases}$$

We do not have an intuitive explanation for the presence of an extra factor of $\frac{1}{2}$ in the boundary cases $\frac{s}{t}=c_1,c_2$.

The results of Theorem 2.2.28 in the concave region S and the boundary lines $\frac{s}{t}=c_1$ or c_2 are heuristically consistent with KPZ type fluctuations. For example, to see this set

$$C=s\,\mathrm{E}\left[\frac{1}{(a+\zeta)^3}\right]+t\,\mathrm{E}\left[\frac{1}{(b-\zeta)^3}\right]=\frac{1}{2}\partial_z^2 g_z(s,t)\big|_{z=\zeta}$$

and assume that our asymptotic result in the concave region hold for finite n. Then for $(s,t) \in S$ and large r, we expect to see

$$\mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor ns \rfloor, \lfloor nt \rfloor) - ng(s,t) \geqslant n^{\frac{1}{3}} C^{\frac{1}{3}} r) \approx \exp\left\{-\frac{4}{3} C^{-\frac{1}{2}} (C^{\frac{1}{3}} n^{-\frac{2}{3}} r)^{\frac{3}{2}} n\right\} = e^{-\frac{4}{3} r^{\frac{3}{2}}},$$

which agrees the leading order large r asymptotics of the Tracy-Widom GUE distribution [4, Exercise 3.8.3]. Note that the choice of normalizing constant C in this argument is not arbitrary. Taking $C = \frac{1}{2}\partial_z^2 g(s,t)|_{z=\zeta}$ is consistent with the normalizing constants needed to see Tracy-Widom GUE limits in, for example, [51, Theorem 1.6] (this is the case $\alpha, \beta \sim \delta_{\frac{1}{2}}$) and in [17, Theorem 1.3]. In the latter case, this was shown to be the constant arising from the KPZ scaling theory in [85]. We also remark that the centering in this argument is likely not correct. As in [38, Theorem 3], we expect that the correct centering should be n times the shape function with α and β given by the empirical distribution of the parameters $\{a_i\}_{i=1}^{\lfloor ns \rfloor}$ and $\{b_j\}_{j=1}^{\lfloor nt \rfloor}$ rather than $ng(s,t)$. This new shape function is not random with respect to $\mathbf{P}_{\mathbf{a},\mathbf{b}}$ and converges to $g(s,t)$ for almost every realization of the environment. Continuity of the rate function then explains why this difference does not appear at the level of right tail large deviations.

Theorem 2.2.29. *Suppose that α and β are not both degenerate. For any $s,t > 0$, as $\epsilon \downarrow 0$,*

$$\mathbb{J}_{s,t}(g(s,t)+\epsilon) = \begin{cases} \left(-s\,\mathrm{E}\left[\frac{1}{a-\underline{\alpha}}\right]^2 + t\,\mathrm{Var}\left[\frac{1}{b+\underline{\alpha}}\right] + t\,\mathrm{E}\left[\frac{1}{(b+\underline{\alpha})^2}\right]\right)^{-1} \epsilon^2/2 + o(\epsilon^2) & \textit{if } s/t < c_1 \\ \left(s\,\mathrm{Var}\left[\frac{1}{a+\zeta}\right] + t\,\mathrm{Var}\left[\frac{1}{b-\zeta}\right]\right)^{-1} \epsilon^2/2 + o(\epsilon^2) & \textit{if } c_1 \leqslant s/t \leqslant c_2 \\ \left(s\,\mathrm{Var}\left[\frac{1}{a+\underline{\beta}}\right] + s\,\mathrm{E}\left[\frac{1}{(a+\underline{\beta})^2}\right] - t\,\mathrm{E}\left[\frac{1}{b-\underline{\beta}}\right]^2\right)^{-1} \epsilon^2/2 + o(\epsilon^2) & \textit{if } s/t > c_2 \end{cases}$$

We do not have any explicitly computable examples for which the regions S_1 and S_2 are non-trivial, but we illustrate the results of the last two theorems with a numerical example.

Example 2.2.30. Choose $\alpha = 4(a-1)^3 1_{[1,2]}(a)da$ and $\beta = \delta_1$. We note that $\underline{\alpha} = \underline{\beta} = 1$. Explicit computation shows that $\mathrm{E}\left[(a-1)^{-2}\right] = 2$, $\mathrm{E}\left[(b-1)^{-2}\right] = \infty$, and $\mathrm{E}\left[(b+1)^{-2}\right] = \frac{1}{4}$. The linear region is then $\frac{s}{t} < \frac{1}{8}$. This is illustrated in Figure 8 below.

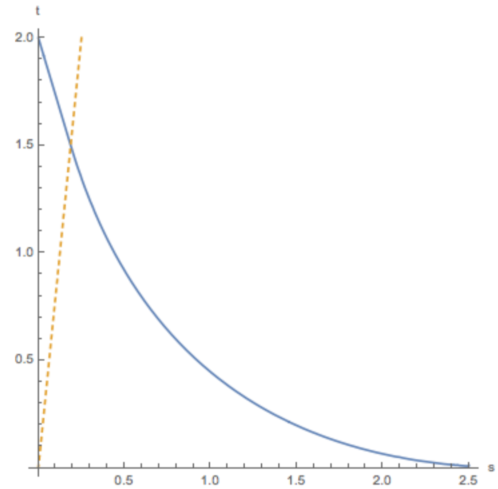
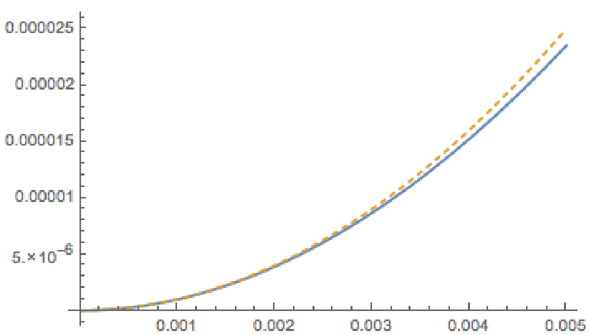
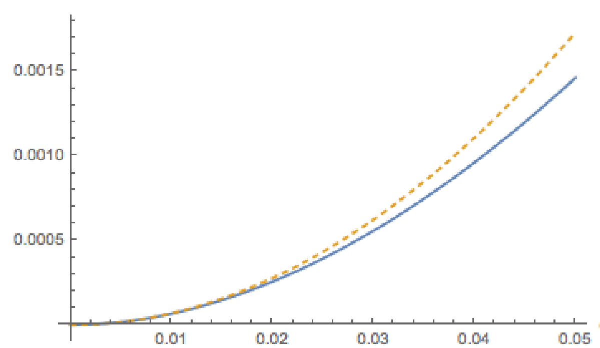


Figure 8: The level set $\{(s,t) : g(s,t) = 1\}$ (solid) and the boundary line $\frac{s}{t} = \frac{1}{8}$ (dashed).

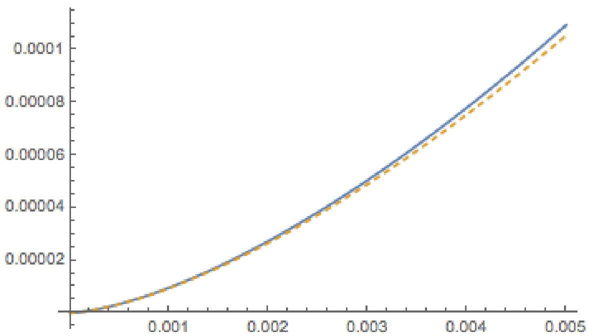
In Figure 9, we plot numerical approximations of the rate functions against the small ϵ asymptotics in Theorems 2.2.28 and 2.2.29. For example, frame (e) plots $\mathbf{J}_{1,1}(g(1,1)+\epsilon)$ against $\frac{4}{3}(E(a+\zeta)^{-3} + \mathrm{E}(b-\zeta)^{-3})^{-\frac{1}{2}}\epsilon^{\frac{3}{2}}$, where ζ is the minimizer in (2.2.14).



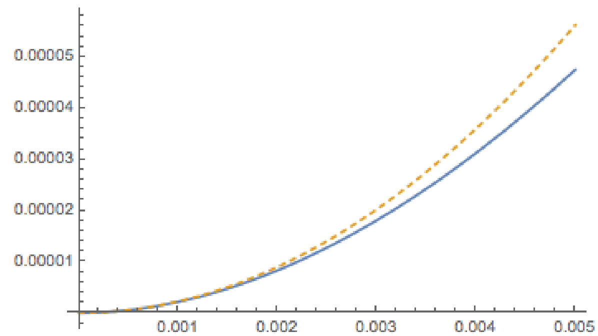
(a) Quenched linear, $t = 10$



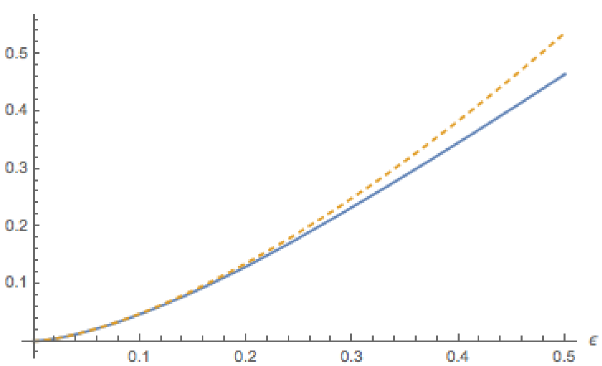
(b) Annealed linear, $t = 10$



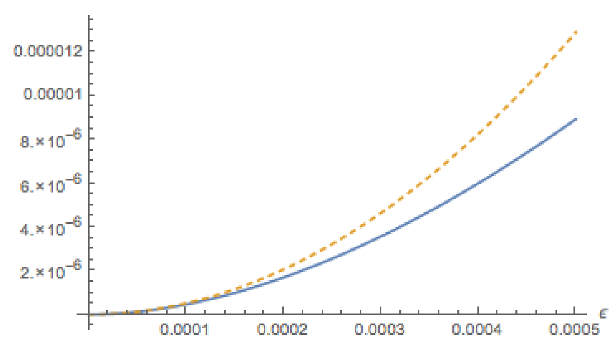
(a) Quenched boundary, $t = 8$



(b) Annealed boundary, $t = 8$



(a) Quenched concave $t = 1$



(b) Annealed concave $t = 1$

Figure 9: Plot of $\mathbb{J}_{s,t}(g(s,t)+\epsilon)$ and $\mathbf{J}_{s,t}(g(s,t)+\epsilon)$ (solid) and their $\epsilon \downarrow 0$ asymptotics (dashed) with $s = 1$.

Stationary model

Extend the space to $\mathbb{R}_+^{\mathbb{Z}_+^2}$. Each weight $W(i,j)$ is now redefined as the projection onto coordinate (i,j) for $i,j \in \mathbb{Z}_+^2$. Introduce the last-passage times

$$\widehat{G}(m,n) = \max_{\pi \in \Pi_{(0,0),(m,n)}} \sum_{i,j \in \pi} W(i,j) \quad \text{for } m,n \in \mathbb{Z}_+\,. \tag{2.2.24}$$

For sequences $\mathbf{a}$ and $\mathbf{b}$ in $(0,\infty)$ and $z \in (-\underline{\alpha}, \underline{\beta})$, define the product measure $\mathbf{P}^z_{\mathbf{a},\mathbf{b}}$ on $\mathbb{R}_+^{\mathbb{Z}_+^2}$ by

$$\begin{aligned} &\mathbf{P}^z_{\mathbf{a},\mathbf{b}}(W(i,j) \geqslant x) = \exp(-(a_i + b_j)x) \qquad &&\mathbf{P}^z_{\mathbf{a},\mathbf{b}}(W(0,0) = 0) = 1 \\ &\mathbf{P}^z_{\mathbf{a},\mathbf{b}}(W(i,0) \geqslant x) = \exp(-(a_i + z)x) \qquad &&\mathbf{P}^z_{\mathbf{a},\mathbf{b}}(W(0,j) \geqslant x) = \exp(-(b_j - z)x) \end{aligned} \tag{2.2.25}$$

for $i,j \in \mathbb{Z}_+$ and $x \geqslant 0$. We will use definition (2.2.25) for $z = -\underline{\alpha}$ when $a_i > \underline{\alpha}$ for $i \in \mathbb{N}$ and for $z = \underline{\beta}$ when $b_j > \underline{\beta}$ for $j \in \mathbb{N}$. The utility of these measures is that the last-passage increments given by $I(m,n) = \widehat{G}(m,n) - \widehat{G}(m-1,n)$ for $m \geqslant 1, n \geqslant 0$ and $J(m,n) = \widehat{G}(m,n) - \widehat{G}(m,n-1)$ for $m \geqslant 0, n \geqslant 1$ are stationary in the following sense.

Proposition 2.2.31 (Proposition 4.1 in [31])**.** *Let $k,l \in \mathbb{Z}_+$. Under $\mathbf{P}^z_{\mathbf{a},\mathbf{b}}$,*

(a) $I(i,l)$ has the same distribution as $W(i,0)$ for $i \in \mathbb{N}$.

(b) $J(k,j)$ has the same distribution as $W(0,j)$ for $j \in \mathbb{N}$.

(c) The random variables $\{I(i,l) : i > k\} \cup \{J(k,j) : j > l\}$ are jointly independent.

For admissible z, define the measure $\mathbb{P}^z$ on $\mathbb{R}_+^{\mathbb{Z}_+^2}$ by $\mathbb{P}^z(B) = \mathrm{E}[\mathbf{P}^z_{\mathbf{a},\mathbf{b}}(B)]$ for any Borel set B. Let $\mathbf{E}^z_{\mathbf{a},\mathbf{b}}$ and $\mathbb{E}^z$ denote the expectations under $\mathbf{P}^z_{\mathbf{a},\mathbf{b}}$ and $\mathbb{P}^z$, respectively. Note from (2.2.25) that the probabilities under $\mathbf{P}^z_{\mathbf{a},\mathbf{b}}$ and $\mathbb{P}^z$ of events generated by $\{W(i,0) : i \in \mathbb{N}\}$ make sense for any $z > -\underline{\alpha}$. Therefore, we permit ourselves to use notation $\mathbf{P}^z_{\mathbf{a},\mathbf{b}}$ and $\mathbb{P}^z$ (and the corresponding expectations) for $z \geqslant \underline{\beta}$ and, similarly, for $z \leqslant -\underline{\alpha}$ when we work only with $\{W(i,0) : i \in \mathbb{N}\}$ and $\{W(0,j) : j \in \mathbb{N}\}$, respectively.

Having proven Theorem 2.2.16, we can also prove the corresponding result for the stationary model with parameter z.

Theorem 2.2.32. *For $z \in [-\underline{\alpha}, \underline{\beta}]$, almost surely for all $s, t > 0$ and $\lambda \in (0, (\underline{\alpha} + z) \wedge (\underline{\beta} - z))$*

$$\mathbf{L}^{z}_{s,t}(\lambda) := \lim_{n\to\infty} n^{-1} \log \mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda \hat{G}(\lfloor ns \rfloor, \lfloor nt \rfloor)}\right]$$
$$= \left\{ s\,\mathrm{E}\left[\log \frac{a+z}{a+z-\lambda}\right] + t\,\mathrm{E}\left[\log \frac{b-z+\lambda}{b-z}\right]\right\} \vee \left\{ s\,\mathrm{E}\left[\log \frac{a+z+\lambda}{a+z}\right] + t\,\mathrm{E}\left[\log \frac{b-z}{b-z-\lambda}\right]\right\}.$$

2.3 Large deviations for Brownian directed percolation

The proofs of the results described above for the Brownian directed percolation model are similar to the proofs of the corresponding results in [32, 37, 49]. The first and last of these papers correspond to the next two sections. Of the models studied in what follows, the proofs are simplest here and so this is the first model we consider.

2.3.1 Proofs for the point-to-point model

As is often the case for directed polymer models, subadditivity plays a key role in the proofs of our large deviation results. In particular, superadditivity of the last passage times in this model gives existence of the right tail rate functions and moment Lyapunov exponents, as shown in the following proposition.

Proposition 2.3.1. *For any $s, t, \lambda > 0$ and $r \in \mathbb{R}$, the limits*

$$\Lambda_{s,t}(\lambda) := \lim_{n\to\infty} \frac{1}{n} \log E\left[e^{\lambda L_{\lfloor ns \rfloor}(nt)}\right], \qquad J_{s,t}(r) := \lim_{n\to\infty} -\frac{1}{n} \log P\left(L_{\lfloor ns \rfloor}(nt) \geqslant nr\right)$$

exist and are real valued. For each $\lambda > 0$, the map $(s,t) \mapsto \Lambda_{s,t}(\lambda)$ for $(s,t) \in (0,\infty)^2$ is positively homogeneous of degree one, superadditive, concave, and continuous. For $(s,t,r) \in (0,\infty)^2 \times \mathbb{R}$, the map $(s,t,r) \mapsto J_{s,t}(r)$ is positively homogeneous of degree one, subadditive, convex, and continuous. For each (s,t), $J_{s,t}(r) = 0$ for $r \leqslant 2\sqrt{st}$ and $r \mapsto J_{s,t}(r)$ is non-decreasing.

Proof. For all of the conclusions except finiteness of $\Lambda_{s,t}(\lambda)$ and the last two properties of $J_{s,t}(r)$, it suffices to show that the maps

$$(s,t) \mapsto \log \mathrm{E}\left[e^{\lambda L_{\lfloor s \rfloor}(t)}\right], \qquad (s,t,r) \mapsto -\log P\left(L_{\lfloor s \rfloor}(t) \geqslant r\right)$$

are superadditive on $(1,\infty) \times (0,\infty)$ and subadditive on $(1,\infty) \times (0,\infty) \times \mathbb{R}$ respectively. See for example the proof of [60, Theorem 16.2.9] and note that a subadditive function which is positively homogeneous of degree one is convex. Take $s_1, s_2 > 1$, $t_1, t_2 > 0$ and $r_1, r_2 \in \mathbb{R}$. We have the inequality

$$L_{\lfloor (s_1+s_2) \rfloor}(t_1 + t_2) \geqslant L_{0,\lfloor s_1 \rfloor}(t_1) + L_{\lfloor s_1 \rfloor, \lfloor (s_1+s_2) \rfloor}(t_1, (t_1 + t_2))$$

where the last two terms are independent. Using translation invariance, independence, and monotonicity of $L_n(t)$ in n, we have

$$\mathrm{E}\left[e^{\lambda L_{\lfloor s_1+s_2 \rfloor}(t_1+t_2)}\right] \geqslant E\left[e^{\lambda L_{\lfloor s_1 \rfloor}(t_1)}\right] E\left[e^{\lambda L_{\lfloor s_2 \rfloor}(t_2)}\right],$$
$$P\left(L_{\lfloor s_1+s_2 \rfloor}(t_1 + t_2) \geqslant r_1 + r_2\right) \geqslant P\left(L_{\lfloor s_1 \rfloor}(t_1) \geqslant r_1\right) P\left(L_{\lfloor s_2 \rfloor}(t_2) \geqslant r_2\right).$$

Finiteness of $\Lambda_{s,t}(\lambda)$ for all $\lambda > 0$ follows from the observation that $L_n(t) \leqslant \sum_{i=0}^{n} 2 \max_{0 \leqslant r \leqslant t} |B_i(r)|$. The properties of $J_{s,t}(r)$ follow from continuity and the fact that the pre-limit expression is non-decreasing in r. □

Remark 2.3.2. Suadditivity shows that $J_{s,t}(r) = \inf_n -n^{-1} \log P\left(L_{\lfloor ns \rfloor}(nt) \geqslant nr\right)$. As a consequence, for any n, we have $P\left(L_{\lfloor ns \rfloor}(nt) \geqslant nr\right) \leqslant \exp\{-nJ_{s,t}(r)\}$.

The next result shows that the decomposition in (2.2.12) implies that $\Lambda_{s,t}(\lambda)$ is the solution to an invertible variational problem. This type of decomposition and versions of the argument that follows are the key steps in the papers [32, 37, 49].

Lemma 2.3.3. *For each $s, t, \lambda > 0$ and any $\mu > \lambda$,*

$$s \log \frac{\mu}{\mu - \lambda} = \sup_{0 \leqslant r < t} \left\{(t - r)\left(\frac{\lambda^2}{2} - \mu\lambda\right) + \Lambda_{s,t-r}(\lambda)\right\}$$
$$\vee \sup_{0 \leqslant u < s} \left\{t\left(\frac{1}{2}\lambda^2 - \mu\lambda\right) + u \log \frac{\mu}{\mu - \lambda} + \Lambda_{s-u,t}(\lambda)\right\}.$$

Proof. We begin with the coupling (2.2.12). It follows that for any $r \in [0, t)$, $u \in [0, s)$ and n large enough,

$$E\left[e^{\lambda \sum_{k=1}^{\lfloor ns \rfloor} q_k^{\mu}(nt)}\right] \geqslant E\left[e^{\lambda(\mu(nr-nt)+B_0(nt)-B_0(nr)+L_{1,n}(nr,nt))}\right]$$

$$\vee\, E\left[e^{\lambda\left(B_0(t)-\mu t+\sum_{k=1}^{\lfloor nu\rfloor} q_k^{\mu}(0)+L_{\lfloor nu\rfloor,n}(0,t)\right)}\right].$$

The random variables $B_0(nt)-B_0(nr)$ and $L_{1,n}(nr,nt)$ are independent because $B_0(\cdot)$ is independent of $\{B_j(\cdot)\}_{j=1}^{\infty}$. The random variables $\sum_{k=1}^{\lfloor nu\rfloor} q_k^{\mu}(0)$, $L_{\lfloor nu\rfloor,n}(0,t)$, and $B_0(t)$ are independent because $\sum_{k=1}^{\lfloor nu\rfloor} q_k^{\mu}(0)$ is measurable with respect to $\sigma(B_j(t): t\leqslant 0, j\in\mathbb{Z}_+)$, $L_{\lfloor nu\rfloor,n}(0,t)$ is measurable with respect to $\sigma(B_j(t): t\geqslant 0, j\in\mathbb{N})$, and $B_0(t)$ is measurable with respect to $\sigma(B_0(t): t>0)$. Taking logs, dividing by n and sending $n\to\infty$, and optimizing over u and r, we immediately obtain $\geqslant$ in the statement of the theorem.

Let $\{r_i\}_{i=1}^{M}$ and $\{u_i\}_{i=1}^{M}$ be uniform partitions of $[0,t]$ and $[0,s]$ respectively. Notice that

$$\begin{aligned}
&\max_{0\leqslant r\leqslant t}\left\{n\mu(r-t)+B_0(nt)-B_0(nr)+L_{1,\lfloor ns\rfloor}(nr,nt)\right\}\\
&=\max_{2\leqslant i\leqslant M}\max_{r\in[r_{i-1},r_i]}\left\{n\mu(r-t)+B_0(nt)-B_0(ns)+L_{1,\lfloor ns\rfloor}(nr,nt)\right\}\\
&\leqslant\max_{2\leqslant i\leqslant M}\left\{n\mu(r_i-t)+B_0(nt)-B_0(nr_i)+\max_{r\in[r_{i-1},r_i]}\{B_0(nr_i)-B_0(nr)\}+L_{1,\lfloor ns\rfloor}(r_{i-1},t)\right\}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\max_{1\leqslant j\leqslant\lfloor ns\rfloor}\left\{B_0(nt)-n\mu t+\sum_{k=1}^{j} q_k^{\mu}(0)+L_{j,\lfloor ns\rfloor}(0,nt)\right\}\\
&\leqslant\max_{2\leqslant i\leqslant M}\left\{B_0(nt)-n\mu t+\sum_{k=1}^{\lfloor nu_i\rfloor} q_k^{\mu}(0)+L_{\lfloor nu_{i-1}\rfloor,\lfloor ns\rfloor}(0,nt)\right\}.
\end{aligned}$$

It follows from these inequalities and independence that

$$\begin{aligned}
&E\left[e^{\lambda\sum_{j=1}^{\lfloor ns\rfloor} q_k^{\mu}(nt)}\right]\\
&\leqslant\sum_{i=2}^{M} e^{n\mu(r_i-t)}E\left[e^{\lambda(B_0(nt)-B_0(nr_i))}\right]E\left[e^{\lambda\max_{r\in[r_{i-1},r_i]}\{B_0(nr_i)-B_0(nr)\}}\right]E\left[e^{\lambda L_{1,\lfloor ns\rfloor}(r_{i-1},t)}\right]\\
&\qquad+E\left[e^{\lambda(B_0(nt)-n\mu t)}\right]E\left[e^{\lambda\sum_{k=1}^{\lfloor nu_i\rfloor} q_k^{\mu}(0)}\right]E\left[e^{\lambda L_{\lfloor nu_{i-1}\rfloor,\lfloor ns\rfloor}(0,nt)}\right]
\end{aligned}$$

By the reflection principle and the assumption that $r_i-r_{i-1}=\frac{1}{M}$, we have

$$E\left[e^{\lambda\max_{r\in[r_{i-1},r_i]} B_0(nr_i)-B_0(nr)}\right]=E\left[e^{\lambda\sqrt{n}|B_0(\frac{1}{M})|}\right]\leqslant E\left[e^{\lambda\sqrt{n}B_0(\frac{1}{M})}\right]+E\left[e^{-\lambda\sqrt{n}B_0(\frac{1}{M})}\right].$$

Take logs, divide by n and send $n \to \infty$ to obtain

$$\begin{aligned} s \log \frac{\mu}{\mu-\lambda} &\leqslant \max_{2 \leqslant i \leqslant M} \left\{ \mu\lambda(r_i - t) + (t - r_i)\frac{\lambda^2}{2} + \frac{\lambda^2}{2M} + \Lambda_{s,t-r_{i-1}}(\lambda) \right\} \\ &\quad \vee \max_{2 \leqslant i \leqslant M} \left\{ \frac{1}{2}\lambda^2 t - \mu\lambda t + u_i \log \frac{\mu}{\mu-\lambda} + \Lambda_{s-u_{i-1},t}(\lambda) \right\} \\ &\leqslant \left(\sup_{0 \leqslant r < t} \left\{ \mu\lambda(r-t) + (t-r)\frac{\lambda^2}{2} + \Lambda_{s,t-r}(\lambda) \right\} + \frac{\lambda^2}{M} + \frac{\mu\lambda}{M} \right) \\ &\quad \vee \left(\sup_{0 \leqslant u < s} \left\{ \frac{1}{2}\lambda^2 t - \mu\lambda t + u \log \frac{\mu}{\mu-\lambda} + \Lambda_{s-u,t}(\lambda) \right\} + \frac{1}{M} \log \frac{\mu}{\mu-\lambda} \right) \end{aligned}$$

Sending $M \to \infty$ completes the proof. □

Variational problems of the type in Lemma 2.3.3 appear for the Lyapunov exponents and time constants (free energies) of directed percolation models (directed polymers) which have associated stationary models that satisfy appropriate analogues of the Burke property. Up to a change of variables, a deformation of the region on which the maximization takes place, and homogeneity of $\Lambda_{s,t}(\lambda)$ in (s,t), this variational expression gives a Legendre-Fenchel duality between directions (s,t) and values of $\mu > \lambda$. See for example [31, Section 5] for this point of view. Alternatively, this variational problem can be solved directly with a bit of calculus. See [49, Proposition 3.10].

Corollary 2.3.4. *For any $s, t, \lambda > 0$,*

$$\begin{aligned} \Lambda_{s,t}(\lambda) &= \min_{\mu > \lambda} \left\{ t\left(\lambda\mu - \frac{1}{2}\lambda^2\right) + s \log \frac{\mu}{\mu-\lambda} \right\} = \min_{z > 0} \left\{ t\left(\frac{1}{2}\lambda^2 + z\lambda\right) + s \log \frac{z+\lambda}{z} \right\} \\ &= \frac{1}{2}\lambda\sqrt{4st + (t\lambda)^2} + s \log\left(\frac{2s + t\lambda^2 + \lambda\sqrt{4st + (t\lambda)^2}}{2s} \right) = \int_0^\lambda \sqrt{4st + (tx)^2}\, dx. \end{aligned}$$

Proof. The first equality follows from Lemma 2.3.3 and [49, Proposition 3.10] (which is Proposition 2.4.10 below) with $I = \{\mu > \lambda\}$, $h(\mu) = -\frac{\lambda^2}{2} + \lambda\mu$, and $g(\mu) = \log \frac{\mu}{\mu-\lambda}$. The second equality is the change of variables $z = \mu - \lambda$. The third and fourth equalities follow from calculus. □

The next result is the analogue of Varadhan's lemma for right tail rate functions. The proof is essentially the same.

Lemma 2.3.5. *For each $s, t > 0$,*

$$\sup_{r\in\mathbb{R}}\{\lambda r - J_{s,t}(r)\} = \begin{cases} \infty & \lambda < 0 \\ \Lambda_{s,t}(\lambda) & \lambda \geqslant 0 \end{cases}.$$

Proof. The result for $\lambda \leqslant 0$ follows from the observations $J_{s,t}(r) \geqslant 0$ for all r, $J_{s,t}(r) = 0$ for $r \leqslant 2\sqrt{st}$ and $r \mapsto J_{s,t}(r)$ is non-decreasing. Take $\lambda, K > 0$, and let $\{m_i\}_{i=1}^M$ be a uniform partition of $[0, K]$. The exponential Markov inequality yields for each $r > 0$

$$\lambda r - J_{s,t}(r) \leqslant \Lambda_{s,t}(\lambda). \tag{2.3.1}$$

Optimizing over r gives $\leqslant$ in the statement of the lemma. For the reverse, notice that

$$\begin{aligned} E\left[e^{\lambda L_{\lfloor ns \rfloor}(nt)}\right] &= \sum_{i=1}^{M} E\left[e^{\lambda L_{\lfloor ns \rfloor}(nt)} 1_{\{L_{\lfloor ns \rfloor}(nt)\in[m_{i-1},m_i)\}}\right] + E\left[e^{\lambda L_{\lfloor ns \rfloor}(nt)} 1_{\{L_{\lfloor ns \rfloor}(nt)\geqslant K\}}\right] \\ &\leqslant \sum_{i=1}^{M} e^{\lambda m_i} P\left(L_{\lfloor ns \rfloor}(nt) \geqslant m_{i-1}\right) + E\left[e^{\lambda L_{\lfloor ns \rfloor}(nt)} 1_{\{L_{\lfloor ns \rfloor}(nt)\geqslant K\}}\right] \\ &\leqslant \sum_{i=1}^{M} e^{\lambda m_i} P\left(L_{\lfloor ns \rfloor}(nt) \geqslant m_{i-1}\right) + E\left[e^{2\lambda L_{\lfloor ns \rfloor}(nt)}\right]^{\frac{1}{2}} P\left(L_{\lfloor ns \rfloor}(nt) \geqslant K\right)^{\frac{1}{2}}. \end{aligned}$$

Take logs, divide by n and send $n \to \infty$ to obtain

$$\begin{aligned} \Lambda_{s,t}(\lambda) &\leqslant \max_{i\leqslant M}\{\lambda m_i - J_{s,t}(m_{i-1})\} \vee \left\{\frac{1}{2}\Lambda_{s,t}(2\lambda) - \frac{1}{2}J_{s,t}(K)\right\} \\ &\leqslant \left(\sup_{r\in\mathbb{R}}\{\lambda r - J_{s,t}(r)\} + \frac{\lambda}{M}\right) \vee \left\{\frac{1}{2}\Lambda_{s,t}(2\lambda) - \frac{1}{2}J_{s,t}(K)\right\}. \end{aligned}$$

Equation (2.3.1) shows that $J_{s,t}(K) \to \infty$ as $K \to \infty$. Sending $M, K \to \infty$ completes the proof. ☐

Corollary 2.3.6. *For $s, t > 0$ and $r \geqslant 2\sqrt{st}$,*

$$J_{s,t}(r) = \sup_{\lambda,z>0}\left\{\lambda r - t\left(\frac{1}{2}\lambda^2 + z\lambda\right) - s\log\frac{z+\lambda}{z}\right\} = \frac{r\sqrt{r^2-4st}}{2t} + s\log\left(\frac{r-\sqrt{r^2-4st}}{r+\sqrt{r^2-4st}}\right).$$

Proof. The first equality follows from Lemma 2.3.5, Corollary 2.3.4, and the Fenchel-Moreau theorem [79, Theorem 12.2]. The second equality can be obtained with calculus. □

Remark 2.3.7. Differentiating the expression in the previous result gives

$$J_{s,t}(r) = \int_0^{r-2\sqrt{st}} t^{-1}\sqrt{x(x+4\sqrt{st})}dx 1_{\{r\geqslant 2\sqrt{st}\}}.$$

Setting $s = t = 1$ and changing variables gives the expression in Theorem 2.2.8. Combining this result with Remark 2.3.2 gives Corollary 2.2.9.

2.3.2 Proofs for the stationary model

Lyapunov exponents

Having computed $\Lambda_{s,t}(\lambda)$, (2.2.11) now leads to a variational problem for the Lyapunov exponents in stationary Brownian directed percolation for each $\mu > \lambda$. As before, this would essentially give the right tail rate function except for the technical point that we no longer have *a priori* existence and convexity of that function. The rate function can be computed directly through an argument parallel to the proof of Lemma 2.3.8 but phrased in terms of right tail rate functions. Note that using Corollary 2.3.4 we may extend $\Lambda_{s,t}(\lambda)$ continuously to $\Lambda_{0,t}(\lambda) = \frac{\lambda^2 t}{2}$ and $\Lambda_{s,0}(\lambda) = 0$.

Lemma 2.3.8. *For each* $\mu, s, t > 0$ *and* $\lambda \in (0, \mu)$,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log E\left[e^{\lambda L^{\mu}_{\lfloor ns\rfloor}(nt)}\right] &= \sup_{0\leqslant r\leqslant t}\left\{r\left(\lambda\mu+\frac{\lambda^2}{2}\right)+\Lambda_{s,t-r}(\lambda)\right\}\vee\sup_{0\leqslant u\leqslant s}\left\{u\log\frac{\mu}{\mu-\lambda}+\Lambda_{s-u,t}(\lambda)\right\}\\ &= \left\{t\left(\frac{\lambda^2}{2}+\mu\lambda\right)+s\log\frac{\mu+\lambda}{\mu}\right\}\vee\left\{t\left(-\frac{\lambda^2}{2}+\mu\lambda\right)+s\log\frac{\mu}{\mu-\lambda}\right\}.\end{aligned}$$

Proof. The proof of the first equality is essentially the same as in the proof of Lemma 2.3.3, except that one must work with lim inf and lim sup. For example, for any $r \in [0, t), u \in [0, s)$, and n sufficiently large, we have

$$E\left[e^{\lambda L^{\mu}_{\lfloor ns \rfloor}(nt)}\right] \geqslant E\left[e^{\lambda(\mu n r - B_0(r))}\right] E\left[e^{\lambda L_{1,\lfloor ns \rfloor}(r,nt)}\right] \vee E\left[e^{\lambda \sum_{i=1}^{\lfloor nu \rfloor} q^{\mu}_k(0)}\right] E\left[e^{\lambda L_{\lfloor nu \rfloor,\lfloor ns \rfloor}(0,nt)}\right].$$

Take logs, divide by n, take $\liminf$, and optimize to obtain

$$\liminf_{n\to\infty} \frac{1}{n} \log E\left[e^{\lambda L^{\mu}_{\lfloor ns \rfloor}(nt)}\right] \geqslant \sup_{0\leqslant r\leqslant t} \left\{ r\left(\lambda\mu + \frac{\lambda^2}{2}\right) + \Lambda_{s,t-r}(\lambda)\right\} \vee \sup_{0\leqslant u\leqslant s} \left\{ u \log \frac{\mu}{\mu-\lambda} + \Lambda_{s-u,t}(\lambda)\right\}.$$

We omit the reverse inequality which is similar. For the second equality, it is convenient to substitute $r \mapsto t-r$ and $u \mapsto s-u$. Using the second variational expression for $\Lambda_{s,r}(\lambda)$ from Corollary 2.3.4 and a minimax theorem (for example, see [77, Appendix B.3]), we obtain

$$\begin{aligned}\sup_{0\leqslant r\leqslant t} \left\{ r\left(\frac{\lambda^2}{2} + \lambda\mu\right) + \Lambda_{s,t-r}(\lambda)\right\} &= t\left(\frac{\lambda^2}{2} + \mu\lambda\right) + \min_{z>0} \max_{0\leqslant r\leqslant t} \left\{ r\lambda(z-\mu) + s\log\frac{z+\lambda}{z}\right\} \\ &= \left\{ t\left(\frac{\lambda^2}{2} + \mu\lambda\right) + s\log\frac{\mu+\lambda}{\mu}\right\} \wedge \min_{z\geqslant\mu} \left\{ t\left(\frac{\lambda^2}{2} + z\lambda\right) + s\log\frac{z+\lambda}{z}\right\}.\end{aligned}$$

The second equality comes from dividing the minimum into the regions $z \leqslant \mu$ and $z > \mu$. A similar argument using the same variational expression and dividing into $z \leqslant \mu - \lambda$ and $z > \mu - \lambda$ shows that

$$\begin{aligned}\sup_{0\leqslant u\leqslant s} \left\{ u\log\frac{\mu}{\mu-\lambda} + \Lambda_{s-u,t}(\lambda)\right\} &= \left\{ t\left(-\frac{\lambda^2}{2} + \mu\lambda\right) + s\log\frac{\mu}{\mu-\lambda}\right\} \\ &\quad \wedge \min_{z\leqslant\mu-\lambda} \left\{ t\left(\frac{\lambda^2}{2} + z\lambda\right) + s\log\frac{z+\lambda}{z}\right\}.\end{aligned}$$

To complete the proof, note that the function being minimized in the second variational expression $\Lambda_{s,r}(\lambda)$ in Corollary 2.3.4 is strictly convex and minimizers exist. □

We complete this section by dealing with exponents $\lambda \geqslant \mu$. This is an immediate corollary of the previous lemma.

Corollary 2.3.9. *For each $\mu, s, t > 0$ and $\lambda \geqslant \mu$,*

$$\lim_{n\to\infty} \frac{1}{n} \log E\left[e^{\lambda L^{\mu}_{\lfloor ns \rfloor}(nt)}\right] = \infty.$$

Proof. The function $\lambda \mapsto \log E\left[e^{\lambda L^{\mu}_{\lfloor ns \rfloor}(nt)}\right]$ is non-decreasing. It follows that for any $\lambda < \mu$,

$$\liminf_{n\to\infty} \frac{1}{n} \log E\left[e^{\mu L^{\mu}_{\lfloor ns \rfloor}(nt)}\right] \geqslant t\left(\frac{\lambda^2}{2} + \lambda\mu\right) + s \log \frac{\mu}{\mu - \lambda}.$$

Sending $\lambda \uparrow \mu$ gives the result. $\square$

Right tail rate functions

We will now work with the rate functions directly. It is convenient to introduce some notation. We will denote the infimal convolution of two functions f and g by $f \mathbin{\square} g(x) := \inf_{y\in\mathbb{R}} \{f(y) + g(x-y)\}$. We will also introduce

$$g^{\mu}_t(x) = \frac{(x-\mu t)^2}{2t} 1_{\{x \geqslant \mu t\}}, \qquad h^{\mu}_s(x) = \left(x\mu - s - s\log\left(\frac{x\mu}{s}\right)\right) 1_{\{x \geqslant \frac{s}{\mu}\}}.$$

which are right tail rate functions for normal and exponential random variables respectively. We will treat the expressions in (2.2.10) separately.

Lemma 2.3.10. *For all $s, t, \mu > 0$ and $x \in \mathbb{R}$,*

$$-n^{-1} \log P\left(\max_{1\leqslant j\leqslant \lfloor ns \rfloor} \left\{L^{\mu}_j(0) + L_{j,\lfloor ns \rfloor}(0, nt)\right\} \geqslant nx\right) = \inf_{0<u<s} \{h^{\mu}_u \mathbin{\square} J_{s-u,t}(x)\}$$

Proof. Using independence and (2.2.11), for each $u \in (0, s)$, and any $x, y \in \mathbb{R}$ we have

$$P\left(\max_{1\leqslant j\leqslant \lfloor ns \rfloor} \left\{L^{\mu}_j(0) + L_{j,\lfloor ns \rfloor}(0, nt)\right\} \geqslant nx\right) \geqslant P\left(L^{\mu}_{\lfloor nu \rfloor}(0) \geqslant ny\right) P\left(L_{\lfloor nu \rfloor,\lfloor ns \rfloor}(0, nt) \geqslant n(x-y)\right).$$

Recall that $L^{\mu}_{\lfloor nu \rfloor}(0) = \sum_{k=1}^{\lfloor nu \rfloor} q^{\mu}_k(0)$ is a sum of i.i.d. exponential random variables with mean μ^{-1}. Take logs, divide by $-n$, and send $n \to \infty$ to obtain

$$\limsup_{n\to\infty} -\frac{1}{n} \log P\left(L^{\mu}_{\lfloor ns \rfloor}(nt) \geqslant nx\right) \leqslant \{h_s(y) + J_{s-u}(x-y)\}$$

We may then optimize over r, u, y obtain $\leqslant$ in the statement of the lemma. For the reverse, we proceed as in Lemma 2.3.3. Take a uniform partition $\{u_i\}_{i=1}^{M}$ and $[0, s]$ respectively. By a union bound, we have

$$P\left(\max_{1\leqslant j\leqslant \lfloor ns \rfloor} \left\{L^{\mu}_j(0) + L_{j,\lfloor ns \rfloor}(0, nt)\right\} \geqslant nx\right) \leqslant \sum_{i=2}^{M} P\left(L^{\mu}_{\lfloor nu_i \rfloor}(0) + L_{\lfloor nu_{i-1} \rfloor,\lfloor ns \rfloor}(0, nt) \geqslant nx\right).$$

Take logs, divide by $-n$ and send $n \to \infty$.

The right tail rate function of an independent sum of random variables becomes the infimal convolution of their right tail rate functions. See for example [37, Lemma 3.6] for a proof of this result which applies in our setting. We obtain

$$\liminf_{n\to\infty} -\frac{1}{n}\log P\left(\max_{1\leqslant j\leqslant \lfloor ns\rfloor}\left\{L_j^{\mu}(0)+L_{j,\lfloor ns\rfloor}(0,nt)\right\}\geqslant nx\right)\geqslant \min_{2\leqslant i\leqslant M}\left\{h^{\mu}_{u_i}\,\square\, J_{s-u_{i-1},t}(x)\right\}.$$

For $\mu, t>0$, $J_{s,t}(y)+h^{\mu}_u(x-y)$ is continuous as a function of $(s,u,y)\in(0,\infty)^2\times\mathbb{R}$ and extends continuously to a function defined on $[0,\infty)^2\times\mathbb{R}$. Note that $J_{s,t}(x)=0$ for $x\leqslant 2\sqrt{st}$ and $h^{\mu}_u(x)=0$ for $x\leqslant u\mu^{-1}$. As a consequence of [37, Lemma 3.6], we may then find a common compact set K so that for all M we have

$$\min_{i\leqslant M}\left\{h^{\mu}_{u_i}\,\square\, J_{s-u_{i-1}}(x)\right\}=\min_{i\leqslant M}\inf_{y\in K}\left\{h^{\mu}_{u_i}(y)+J_{s-u_{i-1},t}(x-y)\right\}$$

Sending $M\to\infty$, this converges to $\inf_{0\leqslant u\leqslant s} h_u \square J_{s-u,t}(x)$ by continuity of the extended function and compactness of $[0,s]^2\times K$. $\square$

Lemma 2.3.11. *For $s,t,\mu>0$ and $x\in\mathbb{R}$,*

$$\inf_{0<u<s}\{h^{\mu}_u\,\square\, J_{s-u,t}(x)\}=\Bigg\{\sup_{0<\lambda<\mu}\max_{0<z\leqslant\mu-\lambda}\left\{\lambda x-t\left(\frac{1}{2}\lambda^2+z\lambda\right)-s\log\frac{z+\lambda}{z}\right\}$$
$$\vee\sup_{0<\lambda<\mu}\left\{\lambda x-t\left(-\frac{1}{2}\lambda^2+\mu\lambda\right)-s\log\frac{\mu}{\mu-\lambda}\right\}\Bigg\}1_{\{x\geqslant \max_{0\leqslant u\leqslant s}\{2\sqrt{t(s-u)}+u\mu^{-1}\}\}}.$$

In particular, when $\mu\leqslant\sqrt{s/t}$,

$$\inf_{0<u<s}\{h^{\mu}_u\,\square\, J_{s-u,t}(x)\}=(2t)^{-1}\int_{t\mu+s\mu^{-1}}^{x}\sqrt{y^2-4st}+(2t\mu-y)dy1_{\{x\geqslant t\mu+s\mu^{-1}\}}.$$

Proof. Note that for $x\leqslant 2\sqrt{t(s-u)}+u\mu^{-1}$, $h^{\mu}_u\square J_{s-u,t}(x)=0$. It therefore suffices to consider $x>\max_{0\leqslant u\leqslant s}\{2\sqrt{t(s-u)}+u\mu^{-1}\}$. There are two cases for the value of this expression: $\mu\leqslant\sqrt{s/t}$ and $\mu\geqslant\sqrt{s/t}$. If $\mu\leqslant\sqrt{s/t}$, then the maximum occurs at $u=s-t\mu^2$ and the maximum is $t\mu+s\mu^{-1}$, which is the time constant in the stationary model. If $\mu\geqslant\sqrt{s/t}$ then the expression being maximized is strictly decreasing and the minimum is $2\sqrt{st}$, which is the time constant in the point-to-point model.

For each $u \in (0,s)$, the functions $h_u^\mu(\cdot), J_{s-u,t}(\cdot)$, and $h_u^\mu \,\square\, J_{s-u,t}(\cdot)$ are non-negative, not infinite, and convex, which implies that they are also proper and continuous. Note further that using Corollary 2.3.4 we may extend $\Lambda_{s,t}(\lambda)$ continuously to $\Lambda_{0,t}(\lambda) = \frac{\lambda^2 t}{2}$ and $\Lambda_{s,0}(\lambda) = 0$ for $s,t,\lambda > 0$. By the Fenchel-Moreau theorem [79, Theorem 12.2],

$$\begin{aligned}
\inf_{0<u<s}\{h_u^\mu \,\square\, J_{s-u,t}(x)\} &= \inf_{0<u<s}\sup_{0<\lambda<\mu}\left\{\lambda x - u\log\frac{\mu}{\mu-\lambda} - \Lambda_{s-u,t}(\lambda)\right\} \\
&= \min_{0\leqslant u\leqslant s}\sup_{0<\lambda<\mu}\left\{\lambda x - u\log\frac{\mu}{\mu-\lambda} - \Lambda_{s-u,t}(\lambda)\right\} \\
&= \sup_{0<\lambda<\mu}\min_{0\leqslant u\leqslant s}\left\{\lambda x - u\log\frac{\mu}{\mu-\lambda} - \Lambda_{s-u,t}(\lambda)\right\} \\
&= \sup_{0<\lambda<\mu}\min_{0\leqslant u\leqslant s}\max_{z>0}\left\{\lambda x - u\log\frac{\mu}{\mu-\lambda} - t\left(\frac{1}{2}\lambda^2 + z\lambda\right) - (s-u)\log\frac{z+\lambda}{z}\right\} \\
&= \sup_{0<\lambda<\mu}\max_{z>0}\min_{0\leqslant u\leqslant s}\left\{\lambda x - u\log\frac{\mu}{\mu-\lambda} - t\left(\frac{1}{2}\lambda^2 + z\lambda\right) - (s-u)\log\frac{z+\lambda}{z}\right\}
\end{aligned}$$

In the second equality we have used continuity of $\Lambda_{s,t}(\lambda)$ at $s = 0$ and in the third and fifth equalities, we have applied a minimax theorem (for example [77, Appendix B.3]). Separating the terms which depend on u from those that do not, the last expression is equal to

$$\sup_{0<\lambda<\mu}\max_{z>0}\left\{\lambda x - t\left(\frac{1}{2}\lambda^2 + z\lambda\right) - s\log\frac{z+\lambda}{z} + \min_{0\leqslant u\leqslant s}\left\{u\left(\log\frac{z+\lambda}{z} - \log\frac{\mu}{\mu-\lambda}\right)\right\}\right\}$$

Next, split the maximum in z into a maximum over $z \leqslant \mu - \lambda$ and $z > \mu - \lambda$. For $z \leqslant \mu - \lambda$, the infimum occurs at $u = 0$ and if $z \geqslant \mu - \lambda$, the minimum occurs at $u = s$. The previous expression is then given by

$$\max_{0<\lambda<\mu}\max_{0<z\leqslant\mu-\lambda}\left\{\lambda x - t\left(\frac{1}{2}\lambda^2 + z\lambda\right) - s\log\frac{z+\lambda}{z}\right\} \vee \max_{0<\lambda<\mu}\left\{\lambda x - t\left(-\frac{1}{2}\lambda^2 + \mu\lambda\right) - s\log\frac{\mu}{\mu-\lambda}\right\}.$$

Note that the left hand side is $\geqslant$ the right hand side of this expression, because for each λ, the term on the right is the value of the function of z evaluated at $z = \mu - \lambda$. For each λ, the two expressions being maximized in λ are equal unless the global extremizer to the maximization problem in z lies in $(0, \mu - \lambda)$. Otherwise, the function in z is strictly increasing on $(0, \mu - \lambda]$ and so the maximum occurs at $\mu - \lambda$. The global minimizer for this function occurs at $\sqrt{s/t + \lambda^2} - \lambda/2$, so the condition that the minimizer lies in $(0, \mu - \lambda)$ is the same as $\sqrt{s/t + \lambda^2} + \lambda/2 < \mu$. This is possible for some value of λ if and only if $\mu > \sqrt{s/t}$.

If $\mu \leqslant \sqrt{s/t}$ then the right and left hand sides are equal and $x > t\mu^{-1} + s\mu$. In this case, the maximum on the right can now be evaluated with calculus:

$$\max_{0<\lambda<\mu} \left\{ \lambda x - t\left(-\frac{1}{2}\lambda^2 + \mu\lambda\right) - s \log \frac{\mu}{\mu - \lambda} \right\} = (2t)^{-1} \int_{t\mu + s\mu^{-1}}^{x} \sqrt{y^2 - 4st} + (2t\mu - y) dy.$$

$\square$

Remark 2.3.12. For $\mu \leqslant \sqrt{s/t}$ and $\epsilon > 0$, the leading order small ϵ asymptotics of the previous expression are

$$(2t)^{-1} \int_{t\mu + s\mu^{-1}}^{t\mu + s\mu^{-1} + \epsilon} \sqrt{y^2 - 4st} + (2t\mu - y) dy = \begin{cases} \frac{2s}{3(st)^{3/4}} \epsilon^{\frac{3}{2}} + o(\epsilon^{\frac{3}{2}}) & \mu = \sqrt{s/t} \\ \frac{\mu^2}{2s - 2t\mu^2} \epsilon^2 + o(\epsilon^2) & \mu < \sqrt{s/t} \end{cases}.$$

Lemma 2.3.13. *For all $s, t, \mu > 0$ and $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} -\frac{1}{n} \log P\left(\max_{0\leqslant r\leqslant t} \left\{ \mu n r - B_0(nr) + L_{1,\lfloor ns \rfloor}(nr, nt) \right\} \geqslant nx \right) = \inf_{0<r<t} \left\{ g_r^\mu \,\square\, J_{s,t-r}(x) \right\}.$$

Proof. As above, the bound

$$\limsup_{n\to\infty} -\frac{1}{n} \log P\left(\max_{0\leqslant r\leqslant t} \left\{ \mu n r - B_0(nr) + L_{1,\lfloor ns \rfloor}(nr, nt) \right\} \geqslant nx \right) \leqslant \inf_{0<r<t} \left\{ g_r^\mu \,\square\, J_{s,t-r}(x) \right\}.$$

follows immediately. The right hand side is zero for $x \leqslant \max_{0\leqslant r\leqslant t}\{\mu r + 2\sqrt{s(t-r)}\}$. By non-negativity of the pre-limit expressions, this implies the result for such x. It then suffices to consider $x > \max_{0\leqslant r\leqslant t}\{\mu r + 2\sqrt{s(t-r)}\}$.

Take a partition $\{r_i\}_{i=1}^M$ of $[0, t]$. Arguing as in Lemma 2.3.3, we have

$$\sum_{i=2}^{M} P\left(n\mu r_i + B_0(nr_i) + \max_{r\in[r_{i-1}, r_i]} (B_0(nr) - B_0(nr_i)) + L_{1,\lfloor ns \rfloor}(nr_{i-1}, nt) \geqslant nx \right)$$

Note that by the reflection principle, the assumption that $r_i - r_{i-1} = M^{-1}$, and Brownian scaling, $\max_{r\in[r_{i-1},r_i]}(B_0(nr) - B_0(nr_i)) \overset{d}{=} \sqrt{n}\sqrt{M}^{-1}|B_0(1)|$. Applying [37, Lemma 3.6], we obtain

$$\liminf_{n\to\infty} -\frac{1}{n}\log P\left(\max_{0\leqslant r\leqslant t}\left\{\mu nr - B_0(nr) + L_{1,\lfloor ns\rfloor}(nr,nt)\right\} \geqslant nx\right) \geqslant \min_{2\leqslant i\leqslant M}\left\{g^{\mu}_{r_i} \,\square\, g^{0}_{M^{-1}} \,\square\, J_{s,t-r_{i-1}}(x)\right\}$$

It is convenient to work with a simpler variational expression than the infimal convolution on the right. All of the functions $g^{\mu}_{r_i}(x), g^{0}_{M^{-1}}(x), J_{s,t-r_{i-1}}(x)$ and $g^{\mu}_{r_i} \,\square\, g^{0}_{M^{-1}} \,\square\, J_{s,t-r_{i-1}}(x)$ are non-negative real valued convex functions and thus continuous and proper. It follows from the Fenchel-Moreau theorem that for $x > \max_{0\leqslant r\leqslant t}\{\mu r + 2\sqrt{s(t-r)}\}$, M sufficiently large, and all i

$$\begin{aligned} g^{\mu}_{r_i} \,\square\, g^{0}_{M^{-1}} \,\square\, J_{s,t-r_{i-1}}(x) &= \sup_{\lambda\geqslant 0}\left\{\lambda x - r_i\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \frac{\lambda^2}{2M} - \Lambda_{s,t-r_{i-1}}(\lambda)\right\} \\ &= \sup_{\lambda\geqslant 0}\left\{\lambda x - r_{i-1}\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \frac{1}{M}\left(\lambda^2 + \mu\lambda\right) - \Lambda_{s,t-r_{i-1}}(\lambda)\right\} \end{aligned}$$

Then we see that for all M sufficiently large,

$$\begin{aligned} \min_{2\leqslant i\leqslant M}\{g^{\mu}_{r_i} \,\square\, g^{0}_{M^{-1}} \,\square\, J_{s,t-r_{i-1}}(x)\} &= \min_{2\leqslant i\leqslant m}\sup_{\lambda>0}\left\{\lambda x - r_{i-1}\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \frac{1}{M}\left(\lambda^2 + \mu\lambda\right) - \Lambda_{s,t-r_{i-1}}(\lambda)\right\} \\ &\geqslant \inf_{0\leqslant r<t}\sup_{\lambda\geqslant 0}\left\{\lambda x - r\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \frac{1}{M}\left(\lambda^2 + \mu\lambda\right) - \Lambda_{s,t-r}(\lambda)\right\} \end{aligned}$$

Explicit computation shows that $\Lambda_{s,t}(\lambda) \geqslant t\lambda^2/2$. Extend $\Lambda_{s,t}(\lambda)$ to $\Lambda_{s,0}(\lambda) = 0$. It follows that for any $r \in [0,t]$,

$$\begin{aligned} \lambda x - r\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \frac{1}{M}\left(\lambda^2 + \mu\lambda\right) - \Lambda_{s,t-r}(\lambda) &\leqslant \lambda x - r\left(\frac{\lambda^2}{2} + \mu\lambda\right) - (t-r)\frac{\lambda^2}{2} \\ &\leqslant \lambda x - t\frac{\lambda^2}{2}. \end{aligned}$$

For any r, if $\lambda > 2xt^{-1}$, the expression inside the supremum is negative. We may therefore restrict the supremum for all r to the set $0 \leqslant \lambda \leqslant 2xt^{-1}$. But now $(r,\lambda) \mapsto \lambda x - r\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \frac{1}{M}\left(\lambda^2 + \mu\lambda\right) - \Lambda_{s,t-r}(\lambda)$ is uniformly continuous on $[0,t] \times [0, 2xt^{-1}]$. Sending $M \to \infty$ gives

$$\liminf_{M\to\infty}\min_{2\leqslant i\leqslant M}\{g^{\mu}_{r_i} \,\square\, g^{0}_{M^{-1}} \,\square\, J_{s,t-r_{i-1}}(x)\} \geqslant \inf_{0\leqslant r<t}\sup_{\lambda\geqslant 0}\left\{\lambda x - r\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \Lambda_{s,t-r}(\lambda)\right\}$$

$$= \inf_{0\leqslant r<t} \{g_r^\mu \,\square\, J_{s,t-r}(x)\} .$$

The result now follows. $\square$

Lemma 2.3.14. *For any $s, t > 0$ and $x \in \mathbb{R}$,*

$$\inf_{0\leqslant r<t} \{g_r^\mu \,\square\, J_{s,t-r}(x)\} = \max_{\lambda\geqslant 0} \max_{z\geqslant \mu} \left\{\lambda x - t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{z+\lambda}{z}\right\}$$
$$\vee \max_{\lambda\geqslant 0} \left\{\lambda x - t\left(\frac{\lambda^2}{2} + \mu\lambda\right) - s\log\frac{\mu+\lambda}{\mu}\right\}$$

In particular, when $\mu \leqslant \sqrt{s/t}$,

$$\inf_{0\leqslant r<t} \{g_r^\mu \,\square\, J_{s,t-r}(x)\} = (2t)^{-1}\int_{2\sqrt{st}}^{x} \left(z - 2t\mu + \sqrt{z^2 - 4st}\right) dz 1_{\{x\geqslant 2\sqrt{st}\}}$$

Proof. Note that $g_r^\mu \square J_{s,t-r}(x) = 0$ for $x \leqslant \max_{0\leqslant r\leqslant t}\{\mu r + 2\sqrt{s(t-r)}\}$. Take $x > \max_{0\leqslant r\leqslant t}\{\mu r + 2\sqrt{s(t-r)}\}$. There are two cases for the value of this maximum. If $\mu \leqslant \sqrt{s/t}$ then $\max_{0\leqslant r\leqslant t}\{\mu r + 2\sqrt{s(t-r)}\} = 2\sqrt{st}$ because the term being maximized is strictly decreasing. If $\mu > \sqrt{s/t}$ then the maximum occurs at $\mu = t + s\mu^{-2}$ and $\max_{0\leqslant r\leqslant t}\{\mu r + 2\sqrt{s(t-r)}\} = t\mu + s\mu^{-1}$. Once again, extend $\Lambda_{s,t}(\lambda)$ to $\Lambda_{s,0}(\lambda) = 0$. We have

$$\begin{aligned}
\inf_{0\leqslant r<t} \{g_r^\mu \,\square\, J_{s,t-r}(x)\} &= \inf_{0\leqslant r<t} \sup_{\lambda\geqslant 0} \left\{\lambda x - r\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \Lambda_{s,t-r}(\lambda)\right\} \\
&= \min_{0\leqslant r\leqslant t} \max_{\lambda\geqslant 0} \left\{\lambda x - r\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \Lambda_{s,t-r}(\lambda)\right\} \\
&= \max_{\lambda\geqslant 0} \min_{0\leqslant r\leqslant t} \left\{\lambda x - r\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \Lambda_{s,t-r}(\lambda)\right\} \\
&= \max_{\lambda\geqslant 0} \min_{0\leqslant r\leqslant t} \max_{z>0} \left\{\lambda x - r\left(\frac{\lambda^2}{2} + \mu\lambda\right) - (t-r)\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{z+\lambda}{z}\right\} \\
&= \max_{\lambda\geqslant 0} \max_{z>0} \min_{0\leqslant r\leqslant t} \left\{\lambda x - r\left(\frac{\lambda^2}{2} + \mu\lambda\right) - (t-r)\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{z+\lambda}{z}\right\}. \\
&= \max_{\lambda\geqslant 0} \max_{z>0} \left\{\lambda x - t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{z+\lambda}{z} + \min_{0\leqslant r\leqslant t}\{r\lambda(z-\mu)\}\right\}.
\end{aligned}$$

We have used minimax theorems in the third and fifth equalities. In order to treat the inner minimum, we separate the maximum in z into the cases $z \leqslant \mu$ or $z \geqslant \mu$. If $z \leqslant \mu$, then the innter minimum occurs at $r = t$. Then we have

$$\begin{aligned}
&\max_{\lambda \geqslant 0} \max_{0 < z \leqslant \mu} \left\{ \lambda x - t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{z+\lambda}{z} + \min_{0 \leqslant r \leqslant t}\{r\lambda(z-\mu)\} \right\} \\
&= \max_{\lambda \geqslant 0} \max_{0 < z \leqslant \mu} \left\{ \lambda x - t\left(\frac{\lambda^2}{2} + \mu\lambda\right) - s\log\frac{z+\lambda}{z} \right\} \\
&= \max_{\lambda \geqslant 0} \left\{ \lambda x - t\left(\frac{\lambda^2}{2} + \mu\lambda\right) - s\log\frac{\mu+\lambda}{\mu} \right\}
\end{aligned}$$

If $z \geqslant \mu$, then the inner minimum occurs at $r = 0$. Combining these cases, we have

$$\begin{aligned}
\inf_{0 \leqslant r < t} \{g_r^\mu \,\square\, J_{s,t-r}(x)\} = &\max_{\lambda \geqslant 0} \max_{z \geqslant \mu} \left\{ \lambda x - t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{z+\lambda}{z} \right\} \\
&\vee \max_{\lambda \geqslant 0} \left\{ \lambda x - t\left(\frac{\lambda^2}{2} + \mu\lambda\right) - s\log\frac{\mu+\lambda}{\mu} \right\}.
\end{aligned}$$

As in the previous case, the second term in the maximum is feasible for the first term by taking $z = \mu$. These two expressions can then only differ if there is a value of λ for which the maximizer of $\lambda x - t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{z+\lambda}{z}$ on $z \in [\mu, \infty)$ occurs in (μ, ∞). If such a z exists, then it occurs at $z = \frac{1}{2}\left(\sqrt{4s/t + \lambda^2} - \lambda\right)$. This function is strictly decreasing in λ, so this can only occur if and only if the value at $\lambda = 0$ is at least μ. This is equivalent to $\mu \leqslant \sqrt{s/t}$. If $\mu \leqslant \sqrt{s/t}$, then the first maximum becomes

$$\begin{aligned}
&\max_{\lambda \geqslant 0} \max_{z \geqslant \mu} \left\{ \lambda x - t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{z+\lambda}{z} \right\} \\
&= \max_{\lambda \geqslant 0} \left\{ x\lambda - \frac{1}{2}\lambda t\sqrt{4s/t + \lambda^2} - 2s \operatorname{arctanh}\left(\frac{\lambda}{\sqrt{4s/t + \lambda^2}}\right) \right\}
\end{aligned}$$

Recalling that $x \geqslant 2\sqrt{st}$ (in general $\mu t + s\mu^{-1} \geqslant 2\sqrt{st}$), the maximum occurs at $\lambda = t^{-1}\sqrt{x^2 - 4st}$. Substituting in, we have

$$\begin{aligned}
\max_{\lambda \geqslant 0} \max_{z \geqslant \mu} \left\{ \lambda x - t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{z+\lambda}{z} \right\} &= t^{-1}\int_{2\sqrt{st}}^{x} \sqrt{z^2 - 4st}\,dz \\
&= J_{s,t}(x).
\end{aligned}$$

For the right hand side, it is convenient to appeal to homogeneity to set $s = 1$. The maximum of $\lambda x - t\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \log\frac{\mu+\lambda}{\mu}$ on $\lambda \geqslant 0$ occurs in $(0, \infty)$ if and only if $x - 2t\mu + \sqrt{x^2 - 4t} > 0$, in which case the maximum occurs at this λ. Otherwise, the maximum occurs at $\lambda = 0$. This function is strictly increasing in x. Thus, this value always lies in the region $(0, \infty)$ if $2\sqrt{t} - 2t\mu > 0$. In particular, this is the case when $\mu < \sqrt{1/t}$. For general s, this condition is $\mu < \sqrt{s/t}$. For $\mu < \sqrt{s/t}$, we are considering an x satisfying $x > 2\sqrt{st}$. We obtain

$$\begin{aligned}\max_{\lambda \geqslant 0} \left\{ \lambda x - t\left(\frac{\lambda^2}{2} + \mu\lambda\right) - s \log \frac{\mu + \lambda}{\mu} \right\} &= s \max_{\lambda \geqslant 0} \left\{ \lambda \frac{x}{s} - \frac{t}{s}\left(\frac{\lambda^2}{2} + \mu\lambda\right) - \log \frac{\mu+\lambda}{\mu} \right\} \\ &= s^2 \int_{2\sqrt{t/s}}^{x/s} (2t)^{-1} \left(y - 2t\mu/s + \sqrt{y^2 - 4t/s} \right) dy \\ &= (2t)^{-1} \int_{2\sqrt{st}}^{x} \left(z - 2t\mu + \sqrt{z^2 - 4st} \right) dz\end{aligned}$$

□

Lemma 2.3.15. *For all $s, t, \mu > 0$ and $x \in \mathbb{R}$, then*

$$J^{\mu}_{s,t}(x) := \lim_{n\to\infty} -n^{-1} \log P\left(L^{\mu}_{\lfloor ns \rfloor}(nt) \geqslant nx \right) = \inf_{0<r<t} \left\{ g^{\mu}_r \,\square\, J_{s,t-r}(x) \right\} \wedge \inf_{0<u<s} \left\{ h^{\mu}_u \,\square\, J_{s-u,t}(x) \right\}.$$

In particular, for $\mu \leqslant \sqrt{s/t}$,

$$J^{\mu}_{s,t}(x) := (2t)^{-1} \int_{t\mu + s\mu^{-1}}^{x} \sqrt{y^2 - 4st} + (2t\mu - y) dy 1_{\{x \geqslant t\mu + s\mu^{-1}\}}$$

Proof. Note that

$$\begin{aligned}&P\left(\sup_{0<r<t} \left\{ n\mu r - B_0(nr) + L_{1,\lfloor ns \rfloor}(nr, nt) \right\} \geqslant nx \right) \vee P\left(\max_{1 \leqslant j \leqslant \lfloor ns \rfloor} \left\{ L^{\mu}_j(0) + L_{j,\lfloor ns \rfloor}(0, nt) \right\} \right) \\ \leqslant &P\left(L^{\mu}_{\lfloor ns \rfloor}(nt) \geqslant nx \right) \\ \leqslant &P\left(\sup_{0<r<t} \left\{ n\mu r - B_0(nr) + L_{1,\lfloor ns \rfloor}(nr, nt) \right\} \geqslant nx \right) + P\left(\max_{1 \leqslant j \leqslant \lfloor ns \rfloor} \left\{ L^{\mu}_j(0) + L_{j,\lfloor ns \rfloor}(0, nt) \right\} \right).\end{aligned}$$

Take logs, divide by $-n$, and send $n \to \infty$ to obtain the first equality. If $\mu \leqslant \sqrt{st}$, this implies that

$$\begin{aligned}J^{\mu}_{s,t}(x) = &\left\{ (2t)^{-1} \int_{t\mu + s\mu^{-1}}^{x} \sqrt{y^2 - 4st} + (2t\mu - y) dy 1_{\{x \geqslant t\mu + s\mu^{-1}\}} \right\} \\ &\wedge \left\{ (2t)^{-1} \int_{2\sqrt{st}}^{x} \left(y - 2t\mu + \sqrt{y^2 - 4st} \right) dy 1_{\{x \geqslant 2\sqrt{st}\}} \right\}\end{aligned}$$

Note that $\mu t + s\mu^{-1} \geqslant 2\sqrt{st}$ for any $\mu > 0$. The first term is then less than or equal to the second for any value of x, from which the result follows. $\square$

2.4 Large deviations for the O'Connell-Yor polymer model

2.4.1 A variational problem for the right tail rate function

Definitions and notation

The goal of this subsection is to introduce the right tail rate function for the free energy, which we will denote $J_{s,t}(x)$, and the rate functions coming from the stationary model which appear in the variational expression for $J_{s,t}(x)$. We will defer some of the proofs of technical results about the existence and regularity of these rate functions to Appendix 2.4.4. We begin by defining these functions and addressing existence.

Theorem 2.4.1. *For all $s \geqslant 0$, $t > 0$ and $r \in \mathbb{R}$, the limit*

$$J_{s,t}(r) = \lim_{n\to\infty} -\frac{1}{n} \log P\left(\log Z_{1,\lfloor ns\rfloor}(0, nt) \geqslant nr\right)$$

exists and is $\mathbb{R}_+$ valued. Moreover, $J_{s,t}(r)$ is continuous, convex, subadditive, and positively homogeneous of degree one as a function of $(s, t, r) \in [0, \infty) \times (0, \infty) \times \mathbb{R}$. For fixed s and t, $J_{s,t}(r)$ is increasing in r and $J_{s,t}(r) = 0$ if $r \leqslant \rho(s, t)$.

The proof of Theorem 2.4.1 can be found in subsection 2.4.4 of Appendix 2.4.4.

Next, we define the computable rate functions from the stationary model. By the Burke property for the stationary model, the first limit below can be computed as the right branch of a Cramér rate function. For $s, t > 0$, we set

$$\begin{aligned} U^\theta_s(x) &= -\lim \frac{1}{n} \log P\left(\sum_{k=1}^{\lfloor ns\rfloor} r^\theta_k(0) \geqslant nx\right) \\ &= \begin{cases} 0 & x \leqslant -s\Psi_0(\theta) \\ x(\theta - \Psi_0^{-1}(-\frac{x}{s})) + s\log\frac{\Gamma(\theta)}{\Gamma(\Psi_0^{-1}(-\frac{x}{s}))} & x > -s\Psi_0(\theta) \end{cases}, \end{aligned}$$

$$R^\theta_t(x) = -\lim \frac{1}{n} \log P\left(B(nt) - \theta nt \geqslant nx\right) = \begin{cases} 0 & x \leqslant -\theta t \\ \frac{1}{2}\left(\frac{x+\theta t}{\sqrt{t}}\right)^2 & x > -\theta t \end{cases}.$$

We may continuously extend $U_s^\theta(x)$ to $s = 0$ by setting

$$U_0^\theta(x) = \begin{cases} 0 & x \leqslant 0 \\ x\theta & x > 0 \end{cases}.$$

We record the Legendre-Fenchel transforms of these functions below:

$$(U_s^\theta)^*(\xi) = \begin{cases} \infty & \xi < 0 \text{ or } \xi \geqslant \theta \\ s \log \frac{\Gamma(\theta-\xi)}{\Gamma(\theta)} & 0 \leqslant \xi < \theta \end{cases}, \qquad (R_t^\theta)^*(\xi) = \begin{cases} \infty & \xi < 0 \\ t(\frac{\xi^2}{2} - \theta\xi) & \xi \geqslant 0 \end{cases}.$$

The next lemma implies existence of the rate functions which will appear when we use equation (2.2.8) to prove that $J_{s,t}(x)$ satisfies a variational problem in the next subsection. Versions of this result appear in several other papers, so we elect not to re-prove it. The exact statement we need appears in [37].

Lemma 2.4.2. *[37, Lemma 3.6]) Suppose that for each n, X_n and Y_n are independent, that the limits*

$$\lambda(s) = \lim_{n\to\infty} -\frac{1}{n} \log P\left(X_n \geqslant ns\right), \qquad \phi(s) = \lim_{n\to\infty} -\frac{1}{n} \log P\left(Y_n \geqslant ns\right)$$

exist, and that λ is continuous. If there exists a_λ and a_ϕ so that $\lambda(a_\lambda) = \phi(a_\phi) = 0$, then

$$\lim_{n\to\infty} -\frac{1}{n} \log P\left(X_n + Y_n \geqslant nr\right) = \begin{cases} \inf_{a_\lambda \leqslant s \leqslant r - a_\phi} \{\phi(r-s) + \lambda(s)\} & r \geqslant a_\phi + a_\lambda \\ 0 & r \leqslant a_\phi + a_\lambda \end{cases}$$

$$= \lambda \,\Box\, \phi(r).$$

We define rate functions corresponding to the two parts of the decomposition in (2.2.8) as follows: for $a \in [0,t)$, $u \in (0,s]$, $v \in [0,s)$, and $x \in \mathbb{R}$ set

$$\begin{aligned} G_{a,s,t}^\theta(x) &= -\lim \frac{1}{n} \log P\left(B(na,nt) - \theta n(t-a) + \log Z_{1,\lfloor ns \rfloor}(na,nt) \geqslant nx\right), \\ H_{u,v,s,t}^\theta(x) &= -\lim \frac{1}{n} \log P\left(-\log Z_0^\theta(nt) + \log Z_{\lfloor nu \rfloor}^\theta(0) + \log Z_{\lfloor nv \rfloor,\lfloor ns \rfloor}(0,nt) \geqslant nx\right). \end{aligned} \tag{2.4.1}$$

Recall that $\log Z_j^\theta(0) = \sum_{k=1}^j r_k^\theta(0)$ is measurable with respect to the sigma algebra $\sigma(B(s), B_k(s) : 1 \leqslant k \leqslant j; s \leqslant 0)$ and that for $0 \leqslant u < nt$, $\log Z_{j,\lfloor ns \rfloor}(u,t)$ is measurable with respect to the sigma algebra $\sigma(B_k(s_k) : j \leqslant k \leqslant \lfloor ns \rfloor, u \leqslant s_j \leqslant nt)$. Combining the independence of the environment with the computations above, Theorem 2.4.1 and Lemma 2.4.2 imply that $G_{a,s,t}^\theta(x)$ and $H_{u,v,s,t}^\theta(x)$ are well-defined. In particular, we immediately obtain

Corollary 2.4.3. *For $a \in [0,t)$ and $u \in (0,s]$, and $v \in [0,s)$*

$$G_{a,s,t}^\theta(x) = R_{t-a}^\theta \,\square\, J_{s,t-a}(x), \qquad H_{u,v,s,t}^\theta(x) = R_t^\theta \,\square\, U_u^\theta \,\square\, J_{s-v,t}(x).$$

In order to show that (2.2.8) leads to a variational problem, we need some regularity on $G_{a,s,t}^\theta(x)$ and $H_{u,v,s,t}^\theta(x)$. The three results that follow are purely technical, so we defer their proofs to subsection 2.4.4 of Appendix 2.4.4. Lemma 2.4.4 gives a strong kind of local uniform continuity of $H_{u,v,s,t}^\theta(x)$ and Lemma 2.4.5 gives the same for $G_{a,s,t}^\theta(x)$. The difference between the two statements comes from Lemma 2.4.6, which shows that $G_{a,s,t}^\theta(x)$ degenerates to infinity locally uniformly near $a = t$.

Lemma 2.4.4. *Fix $\theta, s, t > 0$ and a compact set $K \subseteq \mathbb{R}$. Then*

$$\lim_{\delta,\gamma,\epsilon \downarrow 0} \sup_{\substack{a,b,b' \in [0,s] : |b-b'| < \delta \\ r_1, r_2 \in K : |r_1 - r_2| < \epsilon}} \left\{ |H_{a,b,s,t+\gamma}^\theta(r_1) - H_{a,b',s,t}^\theta(r_2)| \right\} = 0.$$

Lemma 2.4.5. *Fix $\theta, s, t > 0$ and $0 < \delta \leqslant t$ and a compact set $K \subseteq \mathbb{R}$. Then*

$$\lim_{\epsilon,\gamma \downarrow 0} \sup_{\substack{a_1,a_2 \in [0,t-\delta] : |a_1 - a_2| < \gamma \\ r_1, r_2 \in K : |r_1 - r_2| < \epsilon}} \left\{ |G_{a_1,s,t}^\theta(r_1) - G_{a_2,s,t}^\theta(r_2)| \right\} = 0.$$

Lemma 2.4.6. *Fix $\theta, s, t > 0$ and $K \subset \mathbb{R}$ compact. Then*

$$\liminf_{a \uparrow t} \inf_{x \in K} \left\{ G_{a,s,t}^\theta(x) \right\} = \infty.$$

Coarse graining and the variational problem

Fix $a \in [0,t)$ and $0 < \delta \leqslant t-a$. Then (2.2.8) implies the following lower bounds

$$\log\left(n\int_a^{a+\delta}\frac{Z_0^\theta(nu)}{Z_0^\theta(nt)}Z_{1,\lfloor ns\rfloor}(nu,nt)du\right) \leqslant \sum_{k=1}^{\lfloor ns\rfloor} r_k^\theta(nt), \tag{2.4.2}$$

$$-\log Z_0^\theta(nt) + \log Z_j^\theta(0) + \log Z_{j,\lfloor ns\rfloor}(0,nt) \leqslant \sum_{k=1}^{\lfloor ns\rfloor} r_k^\theta(nt). \tag{2.4.3}$$

For any partition $\{a_i\}_{i=0}^N$ of $[0,t]$, we also have

$$\begin{aligned}\sum_{k=1}^{\lfloor ns\rfloor} r_k^\theta(nt) \leqslant \max_{0\leqslant i\leqslant N-1}&\left\{\log\left(n\int_{a_i}^{a_{i+1}}\frac{Z_0^\theta(nu)}{Z_0^\theta(nt)}Z_{1,\lfloor ns\rfloor}(nu,nt)du\right)\right\}\\ &\vee \max_{1\leqslant j\leqslant\lfloor ns\rfloor}\left\{-\log Z_0^\theta(nt)+\log Z_j^\theta(0)+\log Z_{j,\lfloor ns\rfloor}(0,nt)\right\} + \log(N+1+ns).\end{aligned} \tag{2.4.4}$$

Our goal is now to show that estimates (2.4.2), (2.4.3), and (2.4.4) above lead to a variational characterization of the right tail rate function $J_{s,t}(x)$:

$$\begin{aligned}U_s^\theta(x) &= \min\{\inf_{0\leqslant a<t}\left\{G^\theta_{a,s,t}(x)\right\}, \inf_{0\leqslant a<s}\left\{H^\theta_{a,a,s,t}(x)\right\}\}\\ &= \min\{\inf_{0\leqslant a<t}\left\{R^\theta_{t-a}\,\square\, J_{s,t-a}(x)\right\}, \inf_{0\leqslant a<s}\left\{R^\theta_t \,\square\, U^\theta_a \,\square\, J_{s-a,t}(x)\right\}\}.\end{aligned} \tag{2.4.5}$$

To improve the presentation of the paper, we have moved some of the estimates in the proofs that follow to Appendix 2.4.5.

Lemma 2.4.7. *Fix $\theta > 0$, $(s,t) \in (0,\infty)^2$ and $x \in \mathbb{R}$. Then*

$$U_s^\theta(x) \leqslant \min\{\inf_{0\leqslant a<t}\left\{G^\theta_{a,s,t}(x)\right\}, \inf_{0\leqslant a<s}\left\{H^\theta_{a,a,s,t}(x)\right\}\}.$$

Proof. For $a \in [0,s)$, taking $j = \lfloor an\rfloor$ in inequality (2.4.3) above immediately implies

$$U_s^\theta(x) \leqslant H^\theta_{a,a,s,t}(x). \tag{2.4.6}$$

Fix $\delta \in (0,t)$; then for all $a \in [0,t-\delta)$ and all $u \in [0,a+\delta]$, we have

$$Z_{1,1}(nu,n(a+\delta))Z_{1,\lfloor ns\rfloor}(n(a+\delta),nt) \leqslant Z_{1,\lfloor ns\rfloor}(nu,nt). \tag{2.4.7}$$

It then follows that

$$P\left(\log\left(n\int_a^{a+\delta}\frac{Z_0^\theta(nu)}{Z_0^\theta(nt)}Z_{1,\lfloor ns\rfloor}(nu,nt)du\right)\geqslant nx\right)$$
$$\geqslant P\Big(\log Z_{1,\lfloor ns\rfloor}(n(a+\delta),nt)+\log\frac{Z_0^\theta(n(a+\delta))}{Z_0^\theta(nt)}$$
$$+\log\left(n\int_a^{a+\delta}\frac{Z_0^\theta(nu)}{Z_0^\theta(n(a+\delta))}Z_{1,1}(nu,n(a+\delta)du\right)\geqslant nx\Big).$$

Fix $\epsilon>0$. By independence of the Brownian environment, we find that

$$\frac{-1}{n}\log P\left(\log\left(n\int_a^{a+\delta}\frac{Z_0^\theta(nu)}{Z_0^\theta(nt)}Z_{1,\lfloor ns\rfloor}(nu,nt)du\right)\geqslant nx\right)$$
$$\leqslant\frac{-1}{n}\log P\left(\log Z_{1,\lfloor ns\rfloor}(n(a+\delta),nt)+\log\frac{Z_0^\theta(n(a+\delta))}{Z_0^\theta(nt)}\geqslant n(x+\epsilon)\right)\tag{2.4.8}$$
$$+\frac{-1}{n}\log P\left(\log\left(n\int_a^{a+\delta}\frac{Z_0^\theta(nu)}{Z_0^\theta(n(a+\delta))}Z_{1,1}(nu,n(a+\delta)du\right)\geqslant -n\epsilon\right).\tag{2.4.9}$$

Applying the lower bound obtained by considering the minimum of the Brownian increments on the interval $[a,a+\delta]$ allows us to show that as $n\to\infty$ the probability in line (2.4.9) tends to one. Then taking lim sup and recalling inequality (2.4.2), we obtain

$$U_s^\theta(x)\leqslant G^\theta_{a+\delta,s,t}(x+\epsilon).\tag{2.4.10}$$

By Lemma 2.4.5, we may take $\delta,\epsilon\downarrow 0$ in (2.4.10). Optimizing over a in the resulting equation and in (2.4.6) gives the result. □

Lemma 2.4.8. *Fix $\theta>0$, $(s,t)\in(0,\infty)^2$ and $x\in\mathbb{R}$. Then*

$$U_s^\theta(x)\geqslant\min\{\inf_{0\leqslant a<t}\left\{G^\theta_{a,s,t}(x)\right\},\inf_{0\leqslant a<s}\left\{H^\theta_{a,a,s,t}(x)\right\}\}.$$

Proof. Fix a large $p>1$ and small $\epsilon,\gamma>0$. Consider uniform partitions $\{a_i\}_{i=0}^M$ of $[0,t]$ and $\{b_i\}_{i=0}^N$ of $[0,s]$ of mesh $\nu=\frac{t}{M+1}$ and $\delta=\frac{s}{N+1}$ respectively. We will add restrictions on these parameters later in the proof. Take n sufficiently large that $\lfloor b_in\rfloor<\lfloor b_{i+1}n\rfloor$ for all i.

Fix $j<\lfloor ns\rfloor$ not equal to any of the partition points $\lfloor b_in\rfloor$ and consider i so that $\lfloor b_in\rfloor<j<\lfloor b_{i+1}n\rfloor$. Notice that $Z_0^\theta(nt)$ is $\sigma(B(nt))$ measurable and $Z_j^\theta(0)$ is measurable with respect to $\sigma(B(s), B_1(s), \dots B_j(s) : s \leqslant 0)$, so these random variables and $Z_{j,\lfloor ns \rfloor}(u,v)$ are mutually independent if $0 \leqslant u < v$. It follows from translation invariance and this independence that

$$P\left(-\log Z_0^\theta(nt) + \log Z_j^\theta(0) + \log Z_{j,\lfloor ns \rfloor}(0,nt) \geqslant nx\right)$$
$$= P\left(-\log Z_0^\theta(nt) + \log Z_j^\theta(0) + \log Z_{j,\lfloor ns \rfloor}(n\gamma, n(t+\gamma)) \geqslant nx\right).$$

We have

$$Z_{\lfloor b_i n \rfloor, \lfloor ns \rfloor}(0, n(t+\gamma)) \geqslant Z_{\lfloor b_i n \rfloor, j}(0, n\gamma) Z_{j,\lfloor ns \rfloor}(n\gamma, n(t+\gamma)).$$

It then follows that

$$P\left(-\log Z_0^\theta(nt) + \log Z_j^\theta(0) + \log Z_{j,\lfloor ns \rfloor}(0,nt) \geqslant nx\right)$$
$$\leqslant P\left(-\log Z_0^\theta(nt) + \log Z_{\lfloor b_{i+1} n \rfloor}^\theta(0) + \log Z_{\lfloor b_i n \rfloor, \lfloor ns \rfloor}(0, n(t+\gamma)) \geqslant n(x - 2\epsilon)\right)$$
$$+ P\left(\log Z_{\lfloor b_i n \rfloor, j}(0, n\gamma) \leqslant -n\epsilon\right) + P\left(\sum_{k=j+1}^{\lfloor b_{i+1} n \rfloor} r_k^\theta(0) \leqslant -n\epsilon\right).$$

Using the moment bound in Lemma 2.4.21 with $\xi = -p$ for $p > 1$ and the exponential Markov inequality gives the bound

$$P\left(\log Z_{\lfloor b_i n \rfloor, j}(0, n\gamma) \leqslant -n\epsilon\right) \leqslant e^{-np\left(\epsilon - p\gamma - \delta \log \frac{\delta}{\gamma}\right) + o(n)} \leqslant e^{-n\frac{\epsilon}{2}p + o(n)}.$$

For the last inequality, we first require $\gamma < \frac{\epsilon}{4p}$ and then take δ small enough that $\delta \log \frac{\delta}{\gamma} < \frac{\epsilon}{4}$. The exponential Markov inequality and the known moment generating function of the i.i.d. sum give the bound

$$P\left(\sum_{k=j+1}^{\lfloor b_{i+1} n \rfloor} r_k^\theta(0) \leqslant -n\epsilon\right) \leqslant e^{-np\left(\epsilon - \delta p^{-1} \log\left(\Gamma(\theta+p)\Gamma(\theta)^{-1}\right)\right)} \leqslant e^{-np\frac{\epsilon}{2}}$$

where in the last step we additionally require $\delta < \frac{\epsilon p}{4} \log\left(\Gamma(\theta+p)\Gamma(\theta)^{-1}\right)^{-1}$. For the case that j is a partition point, we have

$$P\left(-\log Z_0^\theta(nt) + \log Z_{\lfloor b_i n \rfloor}^\theta(0) + \log Z_{\lfloor b_i n \rfloor, \lfloor ns \rfloor}(0,nt) \geqslant nx\right)$$

$$\leqslant P\left(-\log Z_0^\theta(nt) + \log Z^\theta_{\lfloor b_{i+1}n\rfloor}(0) + Z_{\lfloor b_i n\rfloor,\lfloor ns\rfloor}(0, n(t+\gamma)) \geqslant n(x-2\epsilon)\right)$$
$$+ P\left(\sum_{k=\lfloor b_i n\rfloor}^{\lfloor b_{i+1}n\rfloor} r_k^\theta(0) \leqslant -2n\epsilon\right).$$

and the same error bound as above applies. We now turn to the problem of estimating the integral

$$P\left(\log\left(n\int_{a_i}^{a_{i+1}} \frac{Z_0^\theta(nu)}{Z_0^\theta(nt)} Z_{1,\lfloor ns\rfloor}(nu,nt)du\right) \geqslant nx\right)$$
$$\leqslant P\left(\log\left(\frac{Z_0^\theta(na_i)}{Z_0^\theta(nt)} Z_{1,\lfloor ns\rfloor}(na_i,nt)\right) \geqslant n(x-\epsilon)\right)$$
$$+ P\left(\log\left(n\int_{a_i}^{a_{i+1}} \frac{Z_0^\theta(nu)}{Z_0^\theta(na_i)} \frac{Z_{1,\lfloor ns\rfloor}(nu,nt)}{Z_{1,\lfloor ns\rfloor}(na_i,nt)}du\right) \geqslant n\epsilon\right).$$

We have

$$P\left(\log\left(n\int_{a_i}^{a_{i+1}} \frac{Z_0^\theta(nu)}{Z_0^\theta(na_i)} \frac{Z_{1,\lfloor ns\rfloor}(nu,nt)}{Z_{1,\lfloor ns\rfloor}(na_i,nt)}du\right) \geqslant n\epsilon\right) \leqslant \exp\left\{-n\left(\frac{\epsilon-\theta\nu}{2\sqrt{\nu}}\right)^2 + o(n)\right\}$$

where we require $\nu < \frac{\epsilon}{\theta}$.

Take n sufficiently large that $\log(ns+N) \leqslant n\epsilon$. It follows from (2.4.4) and union bounds that

$$\frac{1}{n}\log P\left(\sum_{k=1}^{\lfloor ns\rfloor} r_k^\theta(nt) \geqslant nx\right) \leqslant \frac{1}{n}\log(ns+N)$$
$$+ \max_{0\leqslant i\leqslant M-1}\left\{\frac{1}{n}\log P\left(\log\left(n\int_{a_i}^{a_{i+1}} Z_0^\theta(nu)Z_0^\theta(nt)^{-1}Z_{1,\lfloor ns\rfloor}(nu,nt)du\right) \geqslant n(x-\epsilon)\right)\right\}$$
$$\vee \max_{1\leqslant j\leqslant \lfloor ns\rfloor}\left\{\frac{1}{n}\log P\left(-\log Z_0^\theta(nt) + \log Z_j^\theta(0) + \log Z_{j,\lfloor ns\rfloor}(0,nt) \geqslant n(x-\epsilon)\right)\right\}.$$

Combining this with the previous estimates, multiplying by -1 and sending $n\to\infty$ gives

$$U_s^\theta(x) \geqslant \min_{0\leqslant i\leqslant M-1}\left\{G^\theta_{a_i s,t}(x-2\epsilon)\right\} \wedge \left(\frac{\epsilon-\theta\nu}{2\sqrt{\nu}}\right)^2 \wedge \frac{p\epsilon}{2} \wedge \min_{0\leqslant i\leqslant N-1}\left\{H_{b_{i+1},b_i,s,t+\gamma}(x-3\epsilon)\right\}$$
$$\geqslant \inf_{a\in[0,t)}\left\{G^\theta_{a,s,t}(x-2\epsilon)\right\} \wedge \left(\frac{\epsilon-\theta\nu}{2\sqrt{\nu}}\right)^2 \wedge \frac{p\epsilon}{2}$$
$$\wedge \inf_{a\in[0,s)}\left\{H_{a,a,s,t}(x) - \sup_{a,b,b'\in[0,s]:|b-b'|<\delta}\left\{|H^\theta_{a,b,s,t+\gamma}(x-3\epsilon) - H^\theta_{a,b',s,t}(x)|\right\}\right\}.$$

We first send $\delta \downarrow 0$, then $\gamma \downarrow 0$, then $\nu \downarrow 0$, then $p \uparrow \infty$. By Lemma 2.4.6, there is $\eta > 0$ so that for all $\epsilon \in [0,1]$, we have

$$\inf_{a\in[0,t)} \left\{ G^{\theta}_{a,s,t}(x-2\epsilon) \right\} = \inf_{a\in[0,t-\eta]} \left\{ G^{\theta}_{a,s,t}(x-2\epsilon) \right\}.$$

Now, take $\epsilon \downarrow 0$ and use Lemmas 2.4.4 and 2.4.5. This gives the desired bound

$$U^{\theta}_s(x) \geqslant \min\{ \inf_{a\in[0,t)} \left\{ G^{\theta}_{a,s,t}(x) \right\}, \inf_{a\in[0,t)} \left\{ H^{\theta}_{a,a,s,t}(x) \right\} \}. \qquad \square$$

We now turn the variational problem for the right tail rate functions into a variational problem involving Legendre-Fenchel transforms.

Lemma 2.4.9. *For any $\theta > 0$ let $\xi \in (0,\theta)$. Then $J^*_{s,t}(\xi)$ satisfies the variational problem*

$$\begin{aligned} 0 = \max\Bigg\{ &\sup_{0\leqslant a<t} \left\{ (t-a)\left(\frac{1}{2}\xi^2 - \theta\xi\right) - s\log\frac{\Gamma(\theta-\xi)}{\Gamma(\theta)} + J^*_{s,t-a}(\xi) \right\}, \\ &\sup_{0\leqslant a<s} \left\{ t\left(\frac{1}{2}\xi^2 - \theta\xi\right) - (s-a)\log\frac{\Gamma(\theta-\xi)}{\Gamma(\theta)} + (J_{s-a,t})^*(\xi) \right\} \Bigg\}. \end{aligned}$$

Proof. Lemma 2.4.7 and Lemma 2.4.8 imply (2.4.5). Infimal convolution is Legendre-Fenchel dual to addition for proper convex functions [79, Theorem 16.4] so we find

$$\begin{aligned} (U^{\theta}_s)^*(\xi) &= \sup_{x\in\mathbb{R}} \left\{ \xi x - \min\left\{ \inf_{0\leqslant a<t} \left\{ R^{\theta}_{t-a} \,\square\, J_{s,t-a}(x) \right\}, \inf_{0\leqslant a<s} \left\{ R^{\theta}_{t} \,\square\, U^{\theta}_a \,\square\, J_{s-a,t}(x) \right\} \right\} \right\} \\ &= \sup_{x\in\mathbb{R}} \left\{ \max\{ \sup_{0\leqslant a<t} \left\{ \xi x - R^{\theta}_{t-a} \,\square\, J_{s,t-a}(x) \right\}, \sup_{0\leqslant a<s} \left\{ \xi x - R^{\theta}_{t} \,\square\, U^{\theta}_a \,\square\, J_{s-a,t}(x) \right\} \right\} \\ &= \max\left\{ \sup_{0\leqslant a<t} \left\{ (R^{\theta}_{t-a})^*(\xi) + J^*_{s,t-a}(\xi) \right\}, \sup_{0\leqslant a<s} \left\{ (R^{\theta}_{t})^*(\xi) + (U^{\theta}_a)^*(\xi) + (J_{s-a,t})^*(\xi) \right\} \right\}. \end{aligned}$$

If $\xi \in (0,\theta)$, then $(U^{\theta}_s)^*(\xi) < \infty$, so we may subtract $(U^{\theta}_s)^*(\xi)$ from both sides. Substituting in the known Legendre-Fenchel transforms gives the result. $\square$

Solving the variational problem

Next, we show that the variational problem in Lemma 2.4.9 identifies $J^*_{s,t}(\xi)$ for $\xi > 0$. To show the analogous result in [37], the authors followed the approach of rephrasing the variational problem as a Legendre-Fenchel transform in the space-time variables and appealing to convex analysis. We present an alternate method for computing $J^*_{s,t}(\xi)$ in the next proposition, which has the advantage of allowing us to avoid some of the technicalities in that argument. This direct approach is the main reason we are able to appeal to the Gärtner-Ellis theorem to prove the large deviation principle.

Proposition 2.4.10. *Let $I \subseteq \mathbb{R}$ be open and connected and let $h, g : I \to \mathbb{R}$ be twice continuously differentiable functions with $h'(\theta) > 0$ and $g'(\theta) < 0$ for all $\theta \in I$. For $(x,y) \in (0,\infty)^2$, define*

$$f_{x,y}(\theta) = xh(\theta) + yg(\theta)$$

and suppose that $\frac{d^2}{d\theta^2} f_{x,y}(\theta) > 0$ for all $(x,y) \in (0,\infty)^2$ and that $f_{x,y}(\theta) \to \infty$ as $\theta \to \partial I$ (which may be a limit as $\theta \to \pm\infty$). If $\Lambda(x,y)$ is a continuous function on $(0,\infty)^2$ with the property that for all $(x,y) \in (0,\infty)^2$ and $\theta \in I$ the identity

$$0 = \sup_{0 \leqslant a < x} \{\Lambda(x-a,y) - f_{x-a,y}(\theta)\} \vee \sup_{0 \leqslant b < y} \{\Lambda(x,y-b) - f_{x,y-b}(\theta)\} \tag{2.4.11}$$

holds, then

$$\Lambda(x,y) = \min_{\theta \in I} \{f_{x,y}(\theta)\}.$$

Proof. Fix $(x,y) \in (0,\infty)^2$ and call $\nu = \frac{y}{x}$. Under these hypotheses, there exists a unique $\theta^*_{x,y} = \arg\min_{\theta \in I} f_{x,y}(\theta) = \theta^*_{1,\nu}$. Identity (2.4.11) implies that for all $a \in [0,x)$ and $b \in [0,y)$ we have

$$\Lambda(x-a,y) \leqslant f_{x-a,y}(\theta^*_{x-a,y}), \qquad \Lambda(x,y-b) \leqslant f_{x,y-b}(\theta^*_{x,y-b}),$$

and therefore for any $\theta \in I$, $a \in [0,x)$ and $b \in [0,y)$,

$$\Lambda(x-a,y) - f_{x-a,y}(\theta) \leqslant f_{x-a,y}(\theta^*_{x-a,y}) - f_{x-a,y}(\theta), \tag{2.4.12}$$

$$\Lambda(x, y-b) - f_{x,y-b}(\theta) \leqslant f_{x,y-b}(\theta^*_{x,y-b}) - f_{x,y-b}(\theta). \tag{2.4.13}$$

Uniqueness of minimizers implies that $f_{x-a,y}(\theta^*_{x-a,y}) - f_{x-a,y}(\theta) < 0$ unless $\theta = \theta^*_{x-a,y}$ and similarly $f_{x,y-b}(\theta^*_{x,y-b}) - f_{x,y-b}(\theta) < 0$ unless $\theta = \theta^*_{x,y-b}$. Notice that $\theta^*_{1,\nu}$ solves

$$0 = h'(\theta^*_{1,\nu}) + \nu g'(\theta^*_{1,\nu}). \tag{2.4.14}$$

By the implicit function theorem, we may differentiate the previous expression with respect to ν to obtain

$$\frac{d\theta^*_{1,\nu}}{d\nu} = -\frac{g'(\theta^*_{1,\nu})}{h''(\theta^*_{1,\nu}) + \nu g''(\theta^*_{1,\nu})} > 0. \tag{2.4.15}$$

Now, set $\theta = \theta^*_{x,y}$ in (2.4.11). Equality (2.4.15) implies that for $a \in (0, x)$ and $b \in (0, y)$, $\theta^*_{(x,y-b)} < \theta^*_{(x,y)} < \theta^*_{(x-a,y)}$. Then (2.4.12) and (2.4.13) give us the inequalities

$$\Lambda(x-a, y) - f_{x-a,y}(\theta^*_{x,y}) \leqslant f_{x-a,y}(\theta^*_{x-a,y}) - f_{x-a,y}(\theta^*_{x,y}) < 0, \tag{2.4.16}$$

$$\Lambda(x, y-b) - f_{x,y-b}(\theta^*_{x,y}) \leqslant f_{x,y-b}(\theta^*_{x,y-b}) - f_{x,y-b}(\theta^*_{x,y}) < 0. \tag{2.4.17}$$

Notice that (2.4.11) implies either there exists $a_n \to a \in [0, x]$ or $b_n \to b \in [0, y]$ so that one of the following hold:

$$\Lambda(x-a_n, y) - f_{x-a_n,y}(\theta^*_{x,y}) \to 0, \qquad \Lambda(x, y-b_n) - f_{x,y-b_n}(\theta^*_{x,y}) \to 0.$$

Our goal is to show that the only possible limits are $a_n \to 0$ or $b_n \to 0$, from which the result follows from continuity. Continuity and inequalities (2.4.16) and (2.4.17) rule out the possibilities $a \in (0, x)$ and $b \in (0, y)$ respectively. It therefore suffices to show that

$$\limsup_{a \to x^-} f_{x-a,y}(\theta^*_{x-a,y}) - f_{x-a,y}(\theta^*_{x,y}) < 0, \tag{2.4.18}$$

$$\limsup_{b \to y^-} f_{x,y-b}(\theta^*_{x,y-b}) - f_{x,y-b}(\theta^*_{x,y}) < 0. \tag{2.4.19}$$

We will only write out the proof of (2.4.18), since the proof of (2.4.19) is similar. For any fixed $a \in (0, x)$, we have

$$f_{x-a,y}(\theta^*_{x-a,y}) - f_{x-a,y}(\theta^*_{x,y}) < 0.$$

It suffices to show that the previous expression is decreasing in a. Differentiating the previous expression and using (2.4.14) and the fact that $\theta^*_{(x,y)} < \theta^*_{(x-a,y)}$, we find

$$\begin{aligned}
&\frac{d}{da}\left((x-a)h(\theta^*_{(x-a,y)}) + yg(\theta^*_{(x-a,y)}) - \left[(x-a)h(\theta^*_{(x,y)}) + yg(\theta^*_{(x,y)})\right]\right) \\
&\qquad = h(\theta^*_{(x,y)}) - h(\theta^*_{(x-a,y)}) < 0.
\end{aligned}$$

□

Corollary 2.4.11. *For all $\xi > 0$,*

$$\begin{aligned}
J^*_{s,t}(\xi) &= \min_{\theta > \xi}\left\{t\left(-\frac{\xi^2}{2} + \theta\xi\right) + s\log\frac{\Gamma(\theta-\xi)}{\Gamma(\theta)}\right\} \\
&= \min_{\mu>0}\left\{t\left(\frac{\xi^2}{2} + \xi\mu\right) - s\log\frac{\Gamma(\mu+\xi)}{\Gamma(\mu)}\right\}.
\end{aligned}$$

Proof. It follows from the variational representation in Lemma 2.4.9 that $J^*_{s,t}(\xi)$ is not infinite for any choice of the parameters $\xi, s, t > 0$. It then follows from Lemma 2.4.19 and [79, Theorem 10.1] that $J^*_{s,t}(\xi)$ is continuous in $(s,t) \in (0,\infty)^2$.

Fix ξ and set $I = \{\theta : \theta > \xi\}$. For $\theta \in I$ and $s, t \in (0, \infty)$, define

$$h(\theta) = -\frac{\xi^2}{2} + \theta\xi, \qquad g(\theta) = \log\frac{\Gamma(\theta-\xi)}{\Gamma(\theta)},$$

$$f_{s,t}(\theta) = sh(\theta) + tg(\theta), \qquad \Lambda(s,t) = J^*_{s,t}(\xi).$$

Lemma 2.4.9 shows that with these definitions $J^*_{s,t}(\xi)$ solves the variational problem in Proposition 2.4.10. Because $\Psi_1(x) > 0$ and $\Psi_2(x) < 0$, we see that for $\theta \in I$

$$g'(\theta) = \Psi_0(\theta-\xi) - \Psi_0(\theta) < 0, \qquad g''(\theta) = \Psi_1(\theta-\xi) - \Psi_1(\theta) > 0.$$

It then follows that $\frac{d^2}{d\theta^2} f_{s,t}(\theta) > 0$. Moreover, since $\log\frac{\Gamma(\theta-\xi)}{\Gamma(\theta)}$ grows like $-\xi\log(\theta)$ at infinity and $-\log(\theta-\xi)$ at ξ, $f_{s,t}(\theta)$ also tends to infinity at the boundary of I and the result follows.

The second equality is the substitution $\mu = \theta - \xi$. □

2.4.2 Moment Lyapunov exponents and the LDP

The next result would be Varadhan's theorem if $J_{s,t}(x)$ were a full rate function, rather than a right tail rate function. The proof is somewhat long and essentially the same as the proof of Varadhan's theorem, so we omit it. Details of a similar argument for the stationary log-gamma polymer can be found in [37, Lemma 5.1]. The exponential moment bound needed for the proof follows from Lemma 2.4.21.

Lemma 2.4.12. *For $\xi > 0$,*

$$J^*_{s,t}(\xi) = \lim_{n\to\infty} \frac{1}{n} \log E\left[e^{\xi \log Z_{1,\lfloor ns\rfloor}(0,nt)}\right]$$

and in particular the limit exists.

Remark 2.4.13. Lemma 2.4.12 shows that $J^*_{s,t}(\xi)$ is the ξ moment Lyapunov exponent for the parabolic Anderson model associated to this polymer. With this in mind, the second formula in the statement of Corollary 2.4.11 above agrees with the conjecture in [16, Appendix A.1].

To see this, we first observe that the partition function we study differs slightly from the partition function $Z_\beta(t,n)$ studied in [16] (defined in equation (3) of that paper). Up to normalization constants both $Z_{0,\lfloor ns\rfloor}(0,nt)$ and $Z_\beta(t,n)$ are conditional expectations of functionals of a Poisson path. The normalization constant for $Z_{0,\lfloor ns\rfloor}(0,nt)$ is given by the Lebesgue measure of the Weyl chamber $A_{\lfloor ns\rfloor+1,nt}$, while the normalization constant for $Z_\beta(t,n)$ is $P_{\pi(0)=0}\left(\pi(t)=n\right)$ where $\pi(\cdot)$ is again a rate one Poisson process. There is a further difference in that [16] adds a pinning potential of strength $\frac{\beta}{2}$ at the origin to the definition of $Z_\beta(t,n)$, which introduces a multiplicative factor of $e^{-\frac{\beta}{2}t}$. Combining these changes and restricting to the parameters studied in [16, Appendix A.1], we have the relation

$$e^{-\frac{n}{2}} \frac{P_{\pi(0)=0}\left(\pi(n)=\lfloor n\nu\rfloor\right)}{|A_{\lfloor n\nu\rfloor+1,n}|} Z_{0,\lfloor n\nu\rfloor}(0,n) = Z_1(n,\lfloor n\nu\rfloor).$$

Since $P_{\pi(0)=0}\left(\pi(n)=\lfloor n\nu\rfloor\right)|A_{\lfloor n\nu\rfloor+1,n}|^{-1}=e^{-n}$, Corollary 2.4.11 and Lemma 2.4.12 then imply that for any $k>0$,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log E\left[Z_1(n,\lfloor n\nu\rfloor)^k\right] &= -\frac{3}{2}k+\min_{z>0}\left\{\frac{k^2}{2}+kz-\nu\log\frac{\Gamma(z+k)}{\Gamma(z)}\right\}\\ &= \min_{z>0}\left\{\frac{k(k-3)}{2}+kz-\nu\log\frac{\Gamma(z+k)}{\Gamma(z)}\right\},\end{aligned}$$

which is the extension of the moment Lyapunov exponent $H_k(z_k^0)$ conjectured in the middle of page 24 of [16].

Our next goal is to show that the left tail large deviations are not relevant at the scale we consider. This proof is based on the proof of [37, Lemma 4.2] which contains a small mistake; as currently phrased, the argument in that paper only works for $s,t\in\mathbb{Q}$. This problem can be fixed by altering the geometry of the proof, but doing this adds some technicalities which can be avoided in the model we study. We will follow an argument similar to the original proof for $s\in\mathbb{Q}$, then show that this implies what we need for all s.

Proposition 2.4.14. *Fix $s,t>0$. For all $\epsilon>0$*

$$\liminf_{n\to\infty}-\frac{1}{n}\log P\left(\log Z_{1,\lfloor ns\rfloor}(0,nt)\leqslant n(\rho(s,t)-\epsilon)\right)=\infty.$$

Proof. First we consider the case $s\in\mathbb{Q}$. There exists $M\in\mathbb{N}$ large enough that $M(s\wedge t)\geqslant 1$ and for all $m\geqslant M$ we have

$$\frac{1}{m}E\log Z_{1,\lfloor ms\rfloor}(0,mt)\geqslant\rho(s,t)-\epsilon.$$

Fix $m\geqslant M$ so that $ms\in\mathbb{N}$. We will denote coordinates in $\mathbb{R}^{\lfloor ns\rfloor-1}$ by $(u_1,\ldots,u_{\lfloor ns\rfloor-1})$. For $a,b,s,t\in(0,\infty)$ and $n,k,l\in\mathbb{N}$, define a family of sets $A_{k,a}^{l,b}\subset\mathbb{R}^{\lfloor ns\rfloor-1}$ by

$$A_{k,a}^{l,b}=\{0<u_1<\cdots<u_{k-1}<a<u_k<\cdots<u_{k+l-1}<a+b<u_{k+l}<\cdots<u_{\lfloor ns\rfloor-1}<nt\}.$$

For $j,k,\in\mathbb{Z}^+$, set

$$A_j^k\equiv A_{j\lfloor ms\rfloor+1,(j+k)mt}^{\lfloor ms\rfloor,mt}.$$

For each n sufficiently large that the expression below is greater than one, define

$$N = \left\lfloor \frac{n}{m} - \lfloor\sqrt{n}\rfloor - 2 \right\rfloor,$$

so that we have

$$(n-2m)t \leqslant (N + \lfloor\sqrt{n}\rfloor + 1)mt \leqslant (n-m)t, \tag{2.4.20}$$

$$\left(\lfloor\sqrt{n}\rfloor + 1\right) ms - 1 \leqslant \lfloor ns\rfloor - N\lfloor ms\rfloor \leqslant \left(\lfloor\sqrt{n}\rfloor + 2\right) ms. \tag{2.4.21}$$

With this choice of N, for $0 \leqslant k \leqslant \lfloor\sqrt{n}\rfloor$ and $0 \leqslant j \leqslant N-1$, A_j^k is nonempty. Then for $0 \leqslant k \leqslant \lfloor\sqrt{n}\rfloor$, define

$$D_k = \cap_{j=0}^{N-1} A_j^k \cap \left\{ u : 0 < u_1 < \cdots < u_{(N+1)ms-1} < t\left(n - \frac{m}{2}\right) < u_{(N+1)ms} < \cdots < u_{\lfloor ns\rfloor -1} < nt \right\}.$$

To simplify the formulas that follow, we introduce the notation $s_j = jms$ and $t_j^k = (j+k)mt$. In words, we can think of D_k as the collection of paths from 0 to nt which traverse the bottom line until t_0^k, then for $0 \leqslant j \leqslant N-1$ move from t_j^k to t_{j+1}^k along the next ms lines. The path then moves from t_N^k to $t\left(n - \frac{m}{2}\right)$ along the next ms lines and finally proceeds to nt along the remaining lines. Observe that $\{D_k\}_{k=0}^{\lfloor\sqrt{n}\rfloor}$ is a pairwise disjoint, non-empty family of sets. With the convention $u_0 = 0$ and $u_{\lfloor ns\rfloor} = nt$, we have the bound

$$Z_{1,\lfloor ns\rfloor}(0, nt) \geqslant \sum_{k=0}^{\lfloor\sqrt{n}\rfloor} \int_{D_k} e^{\sum_{i=1}^{\lfloor ns\rfloor} B_i(u_{i-1}, u_i)} du_1 \ldots u_{\lfloor ns\rfloor -1}.$$

In the integral over D_k, for each $0 \leqslant j \leqslant N$ we add and subtract $B_{s_j}(t_j^k)$ in the exponent. Similarly, add and subtract $B_{s_{N+1}}\left(t\left(n - \frac{m}{2}\right)\right)$. The reason for this step is that this will make the product of integrals coming from the definition of D_k into a product of partition functions, as when we showed supermultiplicativity of the partition function in (**??**). Introduce

$$H_k^n = \inf_{t_N^k = u_0 < u_1 < \cdots < u_{ms-1} < u_{ms} = n\left(t - \frac{m}{2}\right)} \left\{ \sum_{i=0}^{ms-1} B_{s_N + i}(u_{i-1}, u_i) \right\}$$

and observe that $H_0^n \leqslant B_{s_N}(t_N^0, t_N^k) + H_k^n$. Let $C > 0$ be a uniform lower bound in n (recall that m is fixed) on the Lebesgue measure of the Weyl chamber in the definition of $H_{\lfloor\sqrt{n}\rfloor}^n$.

Such a bound exists by (2.4.20). Set $I_n = \max_{t^0_{N-1} \leqslant u \leqslant t^{\lfloor\sqrt{n}\rfloor}_{N-1}} \{B_{s_N}(t^0_{N-1}, u)\}$. We have the lower bound

$$Z_{1,\lfloor ns \rfloor}(0, nt) \geqslant C Z_{s_{N+1},\lfloor ns \rfloor}\left(t\left(n - \frac{m}{2}\right), nt\right) e^{H^n_0 - I_n} \left(\sum_{k=0}^{\lfloor\sqrt{n}\rfloor} e^{B_0(0,t^k_0)} \prod_{j=0}^{N-1} Z_{s_j,s_{j+1}}\left(t^k_j, t^k_{j+1}\right)\right).$$

We therefore have the upper bound

$$\begin{aligned}
&P\left(\log Z_{1,\lfloor ns \rfloor}(0, nt) \leqslant -n(\rho(s,t) - 6\epsilon)\right) \\
&\qquad \leqslant P\left(\log Z_{(N+1)ms,\lfloor ns \rfloor}\left(t\left(n - \frac{m}{2}\right), nt\right) \leqslant -n\epsilon - \log C\right) \\
&\qquad + P\left(\max_{0 \leqslant k \leqslant \lfloor\sqrt{n}\rfloor} \sum_{j=0}^{N-1} \log Z_{s_j+1,s_{j+1}}\left(t^k_j, t^k_{j+1}\right) \leqslant -n(\rho(s,t) - 2\epsilon)\right) \\
&\qquad + P\left(H_0 \leqslant -n\epsilon\right) + P\left(\min_k B_0(t^k_0) \leqslant -n\epsilon\right) + P\left(I_n \geqslant n\epsilon\right).
\end{aligned}$$

It follows from translation invariance, Lemma 2.4.25, and (2.4.21) that

$$P\left(\log Z_{(N+1)ms,\lfloor ns \rfloor}\left(t\left(n - \frac{m}{2}\right), nt\right) \leqslant -n\epsilon - \log \frac{mt}{12}\right) = O\left(e^{-n^{\frac{3}{2}}}\right).$$

We have

$$\begin{aligned}
P&\left(\max_{0 \leqslant k \leqslant \lfloor\sqrt{n}\rfloor} \left\{\sum_{j=0}^{N-1} \log Z_{s_j+1,s_{j+1}}\left(t^k_j, t^k_{j+1}\right)\right\} \leqslant -n(\rho(s,t) - 2\epsilon)\right) \\
&\qquad = P\left(\sum_{j=0}^{N-1} \log Z_{s_j+1,s_{j+1}}\left(t^1_j, t^1_{j+1}\right) \leqslant -n(\rho(s,t) - 2\epsilon)\right)^{\lfloor\sqrt{n}\rfloor} = O\left(e^{-c_1 n^{\frac{3}{2}}}\right)
\end{aligned}$$

for some $c_1 > 0$. The first equality comes from the fact that the terms in the maximum are i.i.d. and the second comes from large deviation estimates for an i.i.d. sum once we recall that $N = \frac{n}{m} + o(n)$.

Recall that by (2.4.20), $n\left(t - \frac{m}{2}\right) - t^0_N = O(\sqrt{n})$. It follows from Lemma 2.4.23 that there exist $c_2, C_2 > 0$ so that

$$P\left(H_0 \leqslant -n\epsilon\right) \leqslant C_2 e^{-c_2 n^{\frac{3}{2}}}.$$

The remaining two terms can be controlled with the reflection principle and are $O\left(e^{-\frac{1}{2}n^{\frac{3}{2}}}\right)$.

Now let s be irrational. For each k, fix $\tilde{s}_k < s$ rational with $e^{-k} < |\tilde{s}_k - s| < 2e^{-k}$ and set $\tilde{t}_k = t - \frac{1}{k}$. Call $\alpha_k = s - \tilde{s}_k$ and $\beta_k = t - \tilde{t}_k = \frac{1}{k}$. Superadditivity gives

$$\begin{aligned}P\left(\log Z_{1,\lfloor ns\rfloor}(0,nt) \leqslant n(\rho(s,t)-\epsilon)\right) &\leqslant P\left(\log Z_{1,\lfloor n\tilde{s}_k\rfloor}(0,n\tilde{t}_k) \leqslant n\left(\rho(\tilde{s}_k,\tilde{t}_k) - \frac{\epsilon}{2}\right)\right)\\ &\quad + P\left(\log Z_{\lfloor n\tilde{s}_k\rfloor,\lfloor ns\rfloor}(n\tilde{t}_k,nt) \leqslant n\left(\rho(s,t) - \rho(\tilde{s}_k,\tilde{t}_k) - \frac{\epsilon}{2}\right)\right).\end{aligned}$$

Since $\tilde{s}_k$ is rational, we have already shown that the first term is negligible. Take k sufficiently large that $\rho(s,t) - \rho(\tilde{s}_k,\tilde{t}_k) - \frac{\epsilon}{2} < -\frac{\epsilon}{4}$. By Lemma 2.4.22, we find

$$\liminf_{n\to\infty} -\frac{1}{n}P\left(\log Z_{1,\lfloor ns\rfloor}(0,nt) \leqslant n(\rho(s,t)-\epsilon)\right) \geqslant \alpha_k J_{GUE}\left(\frac{\frac{\epsilon}{4} - \alpha_k\log\beta_k - \alpha_k + \alpha_k\log\alpha_k}{2\sqrt{\alpha_k\beta_k}}\right).$$

Using formula (2.4.26), $J_{GUE}(r) = 4\int_0^r \sqrt{x(x+2)}dx$, it is not hard to see that as $k \to \infty$, this lower bound tends to infinity. $\square$

Lemma 2.4.15. *Fix $s,t > 0$ and $\xi < 0$. Then*

$$\lim_{n\to\infty}\frac{1}{n}\log E\left[e^{\xi\log Z_{1,\lfloor ns\rfloor}(0,nt)}\right] = \xi\rho(s,t).$$

Proof. Fix $\epsilon > 0$ and small and recall that Lemma 2.4.21 and Jensen's inequality imply that for any $\xi < 0$, $\sup_n\left\{\frac{1}{n}\log E\left[e^{\xi\log Z_{1,\lfloor ns\rfloor}(0,nt)}\right]\right\} < \infty$. The lower bound is immediate from

$$\begin{aligned}\frac{1}{n}\log E\left[e^{\xi\log Z_{1,\lfloor ns\rfloor}(0,nt)}\right] &\geqslant \frac{1}{n}\log E\left[e^{\xi\log Z_{1,\lfloor ns\rfloor}(0,nt)}1_{\{\log Z_{1,\lfloor ns\rfloor}(0,nt)\leqslant n(\rho(s,t)+\epsilon)\}}\right]\\ &\geqslant \xi(\rho(s,t)+\epsilon) + \frac{1}{n}\log P(\log Z_{1,\lfloor ns\rfloor}(0,nt) \leqslant n(\rho(s,t)+\epsilon))\end{aligned}$$

once we recall that $P(\log Z_{1,\lfloor ns\rfloor}(0,nt) \leqslant n(\rho(s,t)+\epsilon)) \to 1$.

For the upper bound, we decompose the expectation as follows

$$\begin{aligned}E\left[e^{\xi\log Z_{1,\lfloor ns\rfloor}(0,nt)}\right] =&E\left[e^{\xi\log Z_{1,\lfloor ns\rfloor}(0,nt)}1_{\{\log Z_{1,\lfloor ns\rfloor}(0,nt)>n(\rho(s,t)-\epsilon)\}}\right]\\ &+ E\left[e^{\xi\log Z_{1,\lfloor ns\rfloor}(0,nt)}1_{\{\log Z_{1,\lfloor ns\rfloor}(0,nt)\leqslant n(\rho(s,t)-\epsilon)\}}\right].\end{aligned}$$

Recalling that $P\left(\log Z_{1,\lfloor ns\rfloor}(0,nt) > n(\rho(s,t)-\epsilon)\right) \to 1$, this leads to

$$\limsup_{n\to\infty}\frac{1}{n}\log E\left[e^{\xi\log Z_{1,\lfloor ns\rfloor}(0,nt)}\right]$$

$$\leqslant \max\left\{\xi(\rho(s,t)-\epsilon), \limsup_{n\to\infty}\frac{1}{n}\log E\left[e^{\xi\log Z_{1,\lfloor ns\rfloor}(0,nt)}1_{\{\log Z_{1,\lfloor ns\rfloor}(0,nt)\leqslant n(\rho(s,t)-\epsilon)\}}\right]\right\}.$$

By Cauchy-Schwarz and Proposition 2.4.14

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log E&\left[e^{\xi\log Z_{1,\lfloor ns\rfloor}(0,nt)}1_{\{\log Z_{1,\lfloor ns\rfloor}(0,nt)\leqslant n(\rho(s,t)-\epsilon)\}}\right]\\ &\leqslant \frac{1}{2}\sup_n\left\{\frac{1}{n}\log E\left[e^{2\xi\log Z_{1,\lfloor ns\rfloor}(0,nt)}\right]\right\}\\ &+\limsup_{n\to\infty}\frac{1}{2n}\log P\left(\log Z_{1,\lfloor ns\rfloor}(0,nt)\leqslant n(\rho(s,t)-\epsilon)\right)=-\infty.\end{aligned}$$ □

Combining the previous results, we are led to the proof of Theorem 2.2.2, from which we immediately deduce Theorem 2.2.3.

Proof of Theorem 2.2.2. Lemmas 2.4.12 and 2.4.15 give the limit for $\xi \neq 0$ and the limit for $\xi = 0$ is zero.

Note that $\Lambda_{s,t}(\xi)$ is differentiable for $\xi < 0$ the left derivative at zero is $\rho(s,t)$. For $\xi > 0$, there is a unique $\mu(\xi)$ solving

$$\Lambda_{s,t}(\xi) = t\left(\frac{\xi^2}{2}+\xi\mu(\xi)\right) - s\log\frac{\Gamma(\mu(\xi)+\xi)}{\Gamma(\mu(\xi))}. \tag{2.4.22}$$

This $\mu(\xi)$ is given by the unique solution to

$$0 = t\xi + s\left(\Psi_0(\mu(\xi)) - \Psi_0(\mu(\xi)+\xi)\right), \tag{2.4.23}$$

which can be rewritten as

$$\frac{1}{\xi}\left(\Psi_0(\mu(\xi)+\xi)-\Psi_0(\mu(\xi))\right) = \frac{t}{s}.$$

By the mean value theorem, there exists $x \in [0,\xi)$ so that

$$\Psi_1^{-1}\left(\frac{t}{s}\right) - x = \mu(\xi).$$

Using this, we see that $\Lambda_{s,t}(\xi)$ is continuous at $\xi = 0$. The implicit function theorem implies that $\mu(\xi)$ is smooth for $\xi > 0$. Differentiating (2.4.22) with respect to ξ and applying (2.4.23), we have

$$\frac{d}{d\xi}\Lambda_{s,t}(\xi) = t\,(\xi + \mu(\xi)) - s\Psi_0(\mu(\xi) + \xi).$$

Substituting in for $\mu(\xi)$, appealing to continuity, and taking $\xi \downarrow 0$ gives

$$\begin{aligned}\lim_{\xi\downarrow 0}\frac{d}{d\xi}\Lambda_{s,t}(\xi) &= t\Psi_1^{-1}\left(\frac{t}{s}\right) - s\Psi_0\left(\Psi_1^{-1}\left(\frac{t}{s}\right)\right)\\ &= \rho(s,t)\end{aligned}$$

which implies differentiability at zero and hence at all ξ. □

Proof of Theorem 2.2.3. The large deviation principle holds by Theorem 2.2.2 and the Gärtner-Ellis theorem [29, Theorem 2.3.6]. □

2.4.3 Stationary Lyapunov exponents

Proof of Theorem 2.2.6. First suppose that $\lambda \in (0,\theta)$. We begin with the coupling

$$Z^\theta_{\lfloor ns\rfloor}(nt) = n\int_0^t Z^\theta_0(nu)Z_{1,\lfloor ns\rfloor}(nu,nt)du + \sum_{j=1}^{\lfloor ns\rfloor} Z^\theta_j(0)Z_{j,\lfloor ns\rfloor}(0,nt).$$

For each $\delta > 0$, each $r \in [0, t-\delta]$ and each $v \in [0,s)$ and n sufficiently large, we have

$$\begin{aligned}Z^\theta_{\lfloor ns\rfloor}(nt) &\geqslant \left(n\int_r^{r+\delta} Z^\theta_0(nu)Z_{1,\lfloor ns\rfloor}(nu,nt)du\right) \vee \left(Z^\theta_{\lfloor nv\rfloor}(0)Z_{\lfloor nv\rfloor,\lfloor ns\rfloor}(0,nt)\right)\\ &\geqslant \left(n\delta Z^\theta_0(nr)Z_{1,\lfloor ns\rfloor}(n(r+\delta),nt)\right) \vee \left(Z^\theta_{\lfloor nv\rfloor}(0)Z_{\lfloor nv\rfloor,\lfloor ns\rfloor}(0,nt)\right).\end{aligned}$$

Similarly, fixing the uniform partition $\{t_i\}_{i=1}^M$ of $[0,t]$ and the uniform partition $\{s_i\}_{i=1}^M$ of $[0,s]$, we have

$$Z^\theta_0(nt) \leqslant \frac{n}{M}\sum_{i=1}^M Z^\theta_0(nt_{i+1})Z_{1,\lfloor ns\rfloor}(nt_i,nt) + \frac{\lfloor ns\rfloor}{M}\sum_{i=1}^M Z^\theta_{\lfloor ns_{i+1}\rfloor}(0)Z_{\lfloor ns_i\rfloor,\lfloor ns\rfloor}(0,nt)$$

It then follows from independence that for any $\delta \in (0,t)$, $r \in [0,t-\delta]$ and $v \in [0,s)$, we have

$$\liminf_{n\to\infty} n^{-1}\log E\left[\left(Z^\theta_{\lfloor ns\rfloor}(nt)\right)^\lambda\right] \geqslant \left\{r\left(\lambda\theta + \frac{\lambda^2}{2}\right) + \Lambda_{s,t-r-\delta}(\lambda)\right\} \vee \left\{v\log\frac{\Gamma(\theta-\lambda)}{\Gamma(\theta)} + \Lambda_{s-v,t}(\lambda)\right\}$$

Take $\delta \downarrow 0$ and then optimize over r and v to obtain a lower bound. For the upper bound, independence and the observation that for $a_i \geqslant 0$ and $\lambda > 0$, $(\sum_{i=1}^M a_i)^\lambda \leqslant M^\lambda \max_i a_i^\lambda \leqslant M^\lambda \sum_i a_i^\lambda$ this implies that

$$E\left[\left(Z^\theta_{\lfloor ns \rfloor}(nt)\right)^\lambda\right] \leqslant (n^\lambda \vee \lfloor ns \rfloor^\lambda) \sum_{i=1}^M \Big\{ E\left[\left(Z_0^\theta(nt_{i+1})\right)^\lambda\right] E\left[(Z_{1,\lfloor ns \rfloor}(nt_i, nt))^\lambda\right] \\ + E\left[\left(\left(Z^\theta_{\lfloor ns_{i+1} \rfloor}(0)\right)^\lambda\right] E\left[(Z_{\lfloor ns_i \rfloor, \lfloor ns \rfloor}(0, nt))\right)^\lambda\right] \Big\}$$

Taking logs, dividing by n and sending $n \to \infty$, we obtain

$$\begin{aligned} \limsup_{n\to\infty} n^{-1} \log E\left[\left(Z^\theta_{\lfloor ns \rfloor}(nt)\right)^\lambda\right] &\leqslant \max_{1\leqslant i \leqslant M} \left\{ t_{i+1}\left(\lambda\theta + \frac{\lambda^2}{2}\right) + \Lambda_{s,t-t_i}(\lambda)\right\} \\ &\quad \vee \left\{ s_{i+1} \log \frac{\Gamma(\theta-\lambda)}{\Gamma(\theta)} + \Lambda_{s-s_i,t}(\lambda)\right\} \\ &\leqslant \left\{ \sup_{0\leqslant r\leqslant t} \left\{ r\left(\lambda\theta + \frac{\lambda^2}{2}\right) + \Lambda_{s,t-r}(\lambda)\right\} + \frac{1}{M}\left(\lambda\theta + \frac{\lambda^2}{2}\right)\right\} \\ &\quad \vee \left\{ \sup_{0\leqslant u\leqslant s} \left\{ u \log \frac{\Gamma(\theta-\lambda)}{\Gamma(\theta)} + \Lambda_{s-u,t}(\lambda)\right\} + \frac{1}{M} \log \frac{\Gamma(\theta-\lambda)}{\Gamma(\theta)}\right\}. \end{aligned}$$

Send $M \to \infty$ and combine with the lower bound to obtain

$$\begin{aligned} \lim_{n\to\infty} n^{-1} \log E\left[\left(Z^\theta_{\lfloor ns \rfloor}(nt)\right)^\lambda\right] &= \sup_{0\leqslant r\leqslant t} \left\{ r\left(\lambda\theta + \frac{\lambda^2}{2}\right) + \Lambda_{s,t-r}(\lambda)\right\} \\ &\quad \vee \sup_{0\leqslant u\leqslant s} \left\{ u \log \frac{\Gamma(\theta-\lambda)}{\Gamma(\theta)} + \Lambda_{s-u,t}(\lambda)\right\}. \end{aligned}$$

Substituting in the variational characterization of $\Lambda_{s,t}(\lambda)$, we have

$$\begin{aligned} \Lambda^\theta_{s,t}(\lambda) = &\sup_{0\leqslant r\leqslant t} \left\{ r\left(\frac{\lambda^2}{2} + \lambda\theta\right) + \min_{z>0}\left\{(t-r)\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{\Gamma(\lambda+z)}{\Gamma(z)}\right\}\right\} \\ &\vee \sup_{0\leqslant u\leqslant s} \left\{ u \log \frac{\Gamma(\theta-\lambda)}{\Gamma(\theta)} + \min_{z>0}\left\{t\left(\frac{\lambda^2}{2} + z\lambda\right) - (s-u)\log\frac{\Gamma(\lambda+z)}{\Gamma(z)}\right\}\right\}. \end{aligned}$$

We may apply a minimax theorem to interchange the supremum and the minimum in both of these terms to obtain

$$\Lambda^\theta_{s,t}(\lambda) = \min_{z>0}\left\{ t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{\Gamma(\lambda+z)}{\Gamma(z)} + \sup_{0\leqslant r\leqslant t}\{r(\theta - z)\}\right\}$$

$$\vee \min_{z>0} \left\{ t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{\Gamma(\lambda+z)}{\Gamma(z)} + \sup_{0\leqslant u\leqslant s}\left\{ u\log\frac{\Gamma(\theta-\lambda)}{\Gamma(\theta)}\frac{\Gamma(\lambda+z)}{\Gamma(z)}\right\}\right\}.$$

For the first term, separate the minimum into minima over $z \in (0,\theta]$ and $z \in [\theta,\infty)$ and for the second, separate it into minima over $z \in (0,\theta-\lambda]$ and $z \in [\theta-\lambda,\infty)$. In the first term, when $z \in (0,\theta]$, the maximum occurs at $r = t$. When $z \in [\theta,\infty)$, the maximum occurs at $r = 0$. It follows that

$$\begin{aligned}\min_{z>0}\left\{ t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{\Gamma(\lambda+z)}{\Gamma(z)} + \sup_{0\leqslant r\leqslant t}\{r(\theta-z)\}\right\} &= \left\{ t\left(\frac{\lambda^2}{2} + \theta\lambda\right) - s\log\frac{\Gamma(\lambda+\theta)}{\Gamma(\theta)}\right\} \\ &\quad \wedge \min_{z\geqslant\theta}\left\{ t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{\Gamma(\lambda+z)}{\Gamma(z)}\right\}.\end{aligned}$$

Similarly, for $z \in (0,\theta-\lambda]$ and $z \in [\theta-\lambda,\infty)$ the observation that the function $z \mapsto \log\Gamma(z+\lambda) - \log\Gamma(z)$ is strictly increasing shows that the inner maximum occurs at $u = 0$ for $z \leqslant \theta-\lambda$ and at $u = s$ for $z \geqslant \theta-\lambda$. It then follows that

$$\begin{aligned}&\min_{z>0}\left\{ t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{\Gamma(\lambda+z)}{\Gamma(z)} + \sup_{0\leqslant u\leqslant s}\left\{ u\log\frac{\Gamma(\theta-\lambda)}{\Gamma(\theta)}\frac{\Gamma(\lambda+z)}{\Gamma(z)}\right\}\right\} \\ &\left\{ t\left(-\frac{\lambda^2}{2} + \theta\lambda\right) - s\log\frac{\Gamma(\theta-\lambda)}{\Gamma(\theta)}\right\} \wedge \min_{z\in(0,\theta-\lambda]}\left\{ t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{\Gamma(\lambda+z)}{\Gamma(z)}\right\}\end{aligned}$$

The result then follows from strict convexity of $z \mapsto t\left(\frac{\lambda^2}{2} + z\lambda\right) - s\log\frac{\Gamma(\lambda+z)}{\Gamma(z)}$. The result for $\lambda \geqslant \theta$ follows from monotonicity as in the proof of Corollary 2.3.9. □

2.4.4 Right tail rate functions

Existence and structure of the right tail rate function

We now turn to the problem of showing the existence and regularity of the right tail rate function for the polymer free energy. Our main goal in this subsection is to prove Theorem 2.4.1. As is typical for right tail large deviations, existence and regularity follow from (almost) subadditivity arguments. Because the partition function degenerates for steps with no time component and we do not restrict attention to integer s, it is necessary to tilt time slightly in this argument. We break the proof of Theorem 2.4.1 into two parts: first we show the result with time tilted, then we show that this change does not matter.

Theorem 2.4.16. *For all $s \geqslant 0$, $t > 0$ and $r \in \mathbb{R}$, the limit*

$$J_{s,t}(r) = \lim_{x\to\infty} -\frac{1}{x}\log P\left(\log Z_{0,\lfloor xs\rfloor}(0, xt-1) \geqslant xr\right)$$

exists and is $\mathbb{R}_+$ valued. Moreover, $J_{s,t}(r)$ is continuous, convex, subadditive, and positively homogeneous of degree one as a function of $(s,t,r) \in [0,\infty)\times(0,\infty)\times\mathbb{R}$. For fixed s and t, $J_{s,t}(r)$ is increasing in r and $J_{s,t}(r) = 0$ if $r \leqslant \rho(s,t)$.

Proof. Define the function $T : [0,\infty)\times(1,\infty)\times\mathbb{R} \to \mathbb{R}_+$ by

$$T(x,y,z) = -\log P\left(\log Z_{0,\lfloor x\rfloor}(0,y-1) \geqslant z\right).$$

Lemma 2.4.20 in the appendix implies that $P\left(\log Z_{0,\lfloor x\rfloor}(0,y-1) \geqslant z\right) \neq 0$ and therefore that this function is well-defined.

Take $(x_1,y_1,z_1),(x_2,y_2,z_2) \in [0,\infty)\times(1,\infty)\times\mathbb{R}$ and call $x_{1,2} = \lfloor x_1+x_2\rfloor - \lfloor x_1\rfloor - \lfloor x_2\rfloor \in \{0,1\}$. By (**??**), we have

$$Z_{0,\lfloor x_1+x_2\rfloor}(0,y_1+y_2-1) \geqslant Z_{0,\lfloor x_1\rfloor}(0,y_1-1)Z_{\lfloor x_1\rfloor,\lfloor x_1+x_2\rfloor}(y_1-1,y_1+y_2-1).$$

Independence and translation invariance then imply

$$\begin{aligned} &P\left(\log Z_{0,\lfloor x_1+x_2\rfloor}(0,y_1+y_2-1) \geqslant z_1+z_2\right) \\ &\qquad\qquad \geqslant P\left(\log Z_{0,\lfloor x_1\rfloor}(0,y_1-1) \geqslant z_1\right) P\left(\log Z_{0,\lfloor x_2\rfloor+x_{1,2}}(0,y_2) \geqslant z_2\right). \end{aligned}$$

If $x_{1,2} = 0$ then, recalling that $\log Z_{\lfloor x_2\rfloor,\lfloor x_2\rfloor}(u,t) = B_{\lfloor x_2\rfloor}(u,t)$, we find

$$\begin{aligned} P\left(\log Z_{0,\lfloor x_2\rfloor}(0,y_2) \geqslant z_2\right) &\geqslant P\left(\log Z_{0,\lfloor x_2\rfloor}(0,y_2-1) \geqslant z_2\right) P\left(\log Z_{\lfloor x_2\rfloor,\lfloor x_2\rfloor}(y_2-1,y_2) \geqslant 0\right) \\ &= \frac{1}{2}P\left(\log Z_{0,\lfloor x_2\rfloor}(0,y_2-1) \geqslant z_2\right). \end{aligned}$$

Similarly, when $x_{1,2} = 1$ we have

$$P\left(\log Z_{0,\lfloor x_2\rfloor+1}(0, y_2) \geqslant z_2\right) \geqslant P\left(\log Z_{0,\lfloor x_2\rfloor}(0, y_2 - 1) \geqslant z_2\right) P\left(\log Z_{0,1}(0, 1) \geqslant 0\right).$$

Setting $C = \max\{\log(2), -\log P(\log Z_{0,1}(0, 1) \geqslant 0)\} < \infty$, we find that

$$T(x_1 + x_2, y_1 + y_2, z_1 + z_2) \leqslant T(x_1, y_1, z_1) + T(x_2, y_2, z_2) + C.$$

$T(x, y, z)$ is therefore subadditive with a bounded correction. Non-negativity and Lemma 2.4.20 imply that $T(x, y, z)$ is bounded for x, y, z in a compact subset of its domain. The proof of [60, Theorem 16.2.9] shows that we may define a function on $[0, \infty) \times (0, \infty) \times \mathbb{R}$ by

$$J_{s,t}(r) = \lim_{x\to\infty} -\frac{1}{x} \log P\left(\log Z_{0,\lfloor xs\rfloor}(0, xt - 1) \geqslant xr\right)$$

and that this function satisfies all of the regularity properties in the statement of the theorem except continuity and monotonicity. Monotonicity in r for fixed s and t follows from monotonicity in the prelimit expression. Convexity and finiteness imply continuity on $(0, \infty)^2 \times \mathbb{R}$ [79, Theorem 10.1]. Moreover, [79, Theorem 10.2] gives upper semicontinuity on all of $[0, \infty) \times (0, \infty) \times \mathbb{R}$. It therefore suffices to show lower semicontinuity at the boundary; namely, we need to show

$$\liminf_{(s',t',r')\to(0,t,r)} J_{s',t'}(r') \geqslant J_{0,t}(r).$$

Fix $(t, r) \in (0, \infty) \times \mathbb{R}$ and a sequence $(s_k, t_k, r_k) \in [0, \infty) \times (0, \infty) \times \mathbb{R}$ with $(s_k, t_k, r_k) \to (0, t, r)$. Recall that $\log Z_{0,0}(0, t) = B_0(t)$, so we may compute with the normal distribution to find $J_{0,t}(r) = \frac{r^2}{2t} 1_{\{r\geqslant 0\}}$. From this we can see that if $s_k = 0$ for all sufficiently large k, we have $J_{s_k,t_k}(r_k) \to J_{0,t}(r)$. We may therefore assume without loss of generality that $s_k > 0$ for all k. First observe that if $r \leqslant 0$, then $J_{0,t}(r) = 0$ and the lower bound follows from non-negativity.

If $r > 0$, we may assume without loss of generality that there exists $c > 0$ with $r_k > c$ for all k. By Lemma 2.4.22 in the appendix, for all sufficiently large k we have

$$J_{s_k,t_k}(r_k) \geqslant s_k J_{GUE}\left(\frac{r_k - s_k \log t_k - s_k + s_k \log s_k}{2\sqrt{t_k s_k}} - 1\right),$$

where $J_{GUE}(r) = 4\int_0^r \sqrt{x(x+2)}dx$. Using this formula and calculus, we find that

$$\lim_{k\to\infty} s_k J_{GUE}\left(\frac{r_k - s_k \log t_k - s_k + s_k \log s_k}{2\sqrt{t_k s_k}} - 1\right) = \frac{r^2}{2t}$$

and therefore continuity follows. Lemma 2.2.1 implies that $J_{s,t}(r) = 0$ for $r \leqslant \rho(s,t)$. ☐

Remark 2.4.17. Note that we only address the spatial boundary in the previous result. The reason for this is that the right tail rate function is not continuous at $t = 0$ for any $s > 0$ and $x \in \mathbb{R}$. To see this, we can use the lower bound for $J_{s,t}(r)$ coming from Lemma 2.4.22. As $t \downarrow 0$, this lower bound tends to infinity.

Lemma 2.4.18. *Fix* $(s,t,r) \in (0,\infty)^2 \times \mathbb{R}$. *For any sequences* $s_n, t_n \in \mathbb{N} \times (0,\infty)$ *with* $\frac{1}{n}(s_n, t_n) \to (s,t)$ *we have*

$$J_{s,t}(r) = \lim_{n\to\infty} -\frac{1}{n} \log P\left(\log Z_{0,s_n}(0,t_n) \geqslant nr\right).$$

Proof. Fix $\epsilon < \min(s,t)$ and positive. We will assume that n is large enough that the following conditions hold:

$$\left\lfloor \left(s - \frac{\epsilon}{2}\right) n \right\rfloor < s_n < \left\lfloor \left(s + \frac{\epsilon}{2}\right) n \right\rfloor, \qquad \left(t - \frac{\epsilon}{2}\right) n < t_n < \left(t + \frac{\epsilon}{2}\right) n - 2.$$

We have

$$Z_{0,s_n}(0,t_n) \geqslant Z_{0,\lfloor(s-\epsilon)n\rfloor}(0,(t-\epsilon)n-1) Z_{\lfloor(s-\epsilon)n\rfloor,s_n}((t-\epsilon)n-1,t_n).$$

It follows that

$$\begin{aligned} &P\left(\log Z_{0,s_n}(0,t_n) \geqslant nr\right) \\ &\quad \geqslant P\left(\log Z_{0,\lfloor(s-\epsilon)n\rfloor}(0,(t-\epsilon)n-1) \geqslant nr\right) P\left(\log Z_{0,s_n-\lfloor(s-\epsilon)n\rfloor}(0,t_n-(t-\epsilon)n+1) \geqslant 0\right). \end{aligned}$$

Call $s(n) = s_n - \lfloor(s-\epsilon)n\rfloor$ and $t(n) = t_n - (t-\epsilon)n + 1$ and divide the interval $(0,t(n))$ into $s(n)$ uniform subintervals. We may bound $Z_{0,s(n)}(0,t(n))$ below by a product of i.i.d. random variables:

$$Z_{0,s(n)}(0,t(n)) \geqslant \prod_{i=1}^{s(n)} Z_{i-1,i}\left((i-1)\frac{t(n)}{s(n)}, i\frac{t(n)}{s(n)}\right).$$

Therefore,

$$P\left(\log Z_{0,s(n)}(0,t(n)) \geqslant 0\right) \geqslant P\left(\log Z_{0,1}\left(0,\frac{t(n)}{s(n)}\right) \geqslant 0\right)^{s(n)}.$$

Notice that $\lim \frac{s(n)}{n} = \epsilon$ and $\lim \frac{t(n)}{n} = \epsilon$, so we may further assume without loss of generality that $\frac{1}{2} < \frac{t(n)}{s(n)} < 2$ for all n. We have

$$Z_{0,1}\left(0,\frac{t(n)}{s(n)}\right) \geqslant Z_{0,1}\left(0,\frac{1}{2}\right) Z_{1,1}\left(\frac{1}{2},\frac{t(n)}{s(n)}\right) = Z_{0,1}\left(0,\frac{1}{2}\right) e^{B_1\left(\frac{1}{2},\frac{t(n)}{s(n)}\right)}$$

so that

$$P\left(\log Z_{0,1}\left(0,\frac{t(n)}{s(n)}\right) \geqslant 0\right) \geqslant \frac{1}{2} P\left(\log Z_{0,1}\left(0,\frac{1}{2}\right) \geqslant 0\right).$$

Therefore for $C = \log(2) - \log P\left(\log Z_{0,1}\left(0,\frac{1}{2}\right) \geqslant 0\right)$ and all $\epsilon < \min(s,t)$ we have

$$\limsup -\frac{1}{n}\log P\left(\log Z_{0,s_n}(0,t_n) \geqslant nr\right) \leqslant J_{s-\epsilon,t-\epsilon}(r) + \epsilon C;$$

sending $\epsilon \downarrow 0$ and applying continuity of the rate function gives one inequality. A similar argument gives the lim inf inequality. $\square$

The next corollary follows from convexity of $J_{s,t}(x)$ in $(s,t,x) \in (0,\infty)^2 \times \mathbb{R}$. Details can be found in the first few lines of the proof of [37, Lemma 4.6].

Lemma 2.4.19. *For all $\xi > 0$, $J^*_{s,t}(\xi)$ is concave as a function of $(s,t) \in (0,\infty)^2$.*

Regularity for the stationary right tail rate functions

Next, we turn to regularity for $H^\theta_{u,v,s,t}(x)$ and $G^\theta_{a,s,t}(x)$, which are defined in (2.4.1). We begin with the proof of Lemma 2.4.4. This result is the only point in the paper where we directly use the continuity up to the boundary in Theorem 2.4.1.

Proof of Lemma 2.4.4. Notice that $\theta t, a\Psi_0(\theta)$, and $\rho(s-b, t+\gamma)$ are bounded for $a, b \in [0, s]$ and $\gamma \in [0, t]$. Using this fact and the formula for $H^{\theta}_{a,b,s,t+\gamma}(r)$ coming from Corollary 2.4.3 and Lemma 2.4.2, there exists a compact set K' containing K so that for all $r \in K$, $a, b \in [0, s]$, and $\gamma \in [0, t]$

$$H^{\theta}_{a,b,s,t+\gamma}(r) = \inf_{x \in K'} \{R^{\theta}_t \boxdot U^{\theta}_a(x) + J_{s-b,t+\gamma}(r-x)\}.$$

Note that $(a, x) \mapsto R^{\theta}_t \boxdot U^{\theta}_a(x)$ is continuous on $[0, s] \times \mathbb{R}$. By Theorem 2.4.1, for any compact set K' we have joint uniform continuity of $(a, b, \gamma, r, x) \mapsto R^{\theta}_t \boxdot U^{\theta}_a(x) + J_{b,t+\gamma}(r-x)$ on the compact set $[0, s]^2 \times [0, t] \times K' \times K'$ and so the result follows. □

The proof of Lemma 2.4.5 is similar to the proof of Lemma 2.4.4, so we omit it. Next, we turn to the proof of Lemma 2.4.6, which shows that $G^{\theta}_{a,s,t}(x)$ tends to infinity locally uniformly near near $a = t$.

Proof of Lemma 2.4.6. We have

$$G^{\theta}_{a,s,t}(x) = \begin{cases} 0 & x \leqslant -\theta(t-a) + \rho(s, t-a) \\ \inf\limits_{-\theta(t-a) \leqslant y \leqslant x - \rho(s,t-a)} \{J_{s,t-a}(x-y) + R^{\theta}_{t-a}(y)\} & x > -\theta(t-a) + \rho(s, t-a) \end{cases}.$$

Fix $\epsilon > 0$. The formula in Lemma 2.2.1 shows that $\rho(s, t-a) \to -\infty$ as $a \uparrow t$, so that for all $x \in \mathbb{R}$ and a sufficiently close to $t, x > -\theta(t-a) + \rho(s, t-a)$. For a sufficiently large that this holds for all $x \in K$, we have

$$J_{s,t-a}(x-y) + R_{t-a}(y) \geqslant \begin{cases} J_{s,t-a}(x + \theta(t-a) - \epsilon)) & y \in [-\theta(t-a), -\theta(t-a) + \epsilon] \\ R^{\theta}_{t-a}(-\theta(t-a) + \epsilon) & y \in [-\theta(t-a) + \epsilon, x - \rho(s, t-a) - \epsilon] \\ R^{\theta}_{t-a}(x - \rho(s, t-a) - \epsilon) & y \in [x - \rho(s, t-a) - \epsilon, x - \rho(s, t)] \end{cases}.$$

By Lemma 2.4.22, for all $x \in K$ and a sufficiently large, we have

$$J_{s,t-a}(x) \geqslant sJ_{GUE}\left(\frac{x - s\log(t-a) - s(1 - \log(s))}{2\sqrt{(t-a)s}} - 1\right).$$

Combining this with the exact formula for $R^\theta_{t-a}(x)$ and optimizing the lower bounds over $x \in K$ shows that the infimum over $x \in K$ of the minimum of these three lower bounds tends to infinity, giving the result. □

2.4.5 Technical estimates

To reduce the clutter elsewhere in the paper, we collect a number of useful estimates in this appendix.

A lower bound on the probability of being large

Lemma 2.4.20. *Let $K \subset [0,\infty) \times (0,\infty) \times \mathbb{R}$ be compact. Then there exists $C_K > 0$ so that for all $(x,y,z) \in K$*

$$P(\log Z_{0,\lfloor x \rfloor}(0,y) \geqslant z) \geqslant C_K.$$

Proof. Since $\lfloor x \rfloor$ takes only finitely many values on any compact set, we may fix $\lfloor x \rfloor$. If $\lfloor x \rfloor = 0$, then $Z_{0,\lfloor x \rfloor}(0,y) = B_0(y)$ and the result follows. For $\lfloor x \rfloor \geqslant 1$, we bound below by an i.i.d. product: $Z_{0,\lfloor x \rfloor}(0,y) \geqslant \prod_{i=0}^{\lfloor x \rfloor - 1} Z_{i,i+1}\left(i\frac{y}{\lfloor x \rfloor}, (i+1)\frac{y}{\lfloor x \rfloor}\right)$. It follows that

$$P\left(\log Z_{0,\lfloor x \rfloor}(0,y) \geqslant z\right) \geqslant P\left(\log Z_{0,1}\left(0, \frac{y}{\lfloor x \rfloor}\right) \geqslant \frac{z}{\lfloor x \rfloor}\right)^{\lfloor x \rfloor}. \tag{2.4.24}$$

Jensen's inequality applied to $\log Z_{0,1}(0,t)$ gives

$$\log Z_{0,1}(0,t) = \log \int_0^t e^{B_0(u)+B_1(u,t)} du \geqslant \log(t) + \frac{1}{t}\int_0^t B_0(u)du + \frac{1}{t}\int_0^t B_1(u,t)du, \tag{2.4.25}$$

where $\frac{1}{t}\int_0^t B_0(u)du$ and $\frac{1}{t}\int_0^t B_1(u,t)du$ are i.i.d. mean zero normal random variables with variance $\frac{t}{3}$. Applying this lower bound to the expression in (2.4.24) gives the result. □

Moment estimate for the partition function

Lemma 2.4.21. *Fix $t > 0$, $n \in \mathbb{N}$ and $\xi \in \mathbb{R}$ with $|\xi| > 1$. Then there is a constant $C > 0$ depending only on ξ so that*

$$E\left[Z_{1,n}(0,t)^{\xi}\right] \leqslant C\left(\frac{\sqrt{n}}{t}\left(\frac{te}{n}\right)^{n}\right)^{\xi} e^{\frac{1}{2}\xi^2 t}.$$

Proof. By Jensen's inequality with respect to the uniform measure on $A_{n,t}$ and Tonelli's theorem we find

$$\begin{aligned} E\left[Z_{1,n}(0,t)^{\xi}\right] &= E\left[\left(\int_{0<s_1<\cdots<s_{n-1}<t} e^{\sum_i B(s_i,s_{i+1})} ds_1 \dots ds_{n-1}\right)^{\xi}\right] \\ &\leqslant |A_{n,t}|^{-1}|A_{n,t}|^{\xi}\int_{0<s_1<\cdots<s_{n-1}<t} E[e^{\xi\sum_i B(s_i,s_{i+1})}]ds_1\dots ds_{n-1} \\ &= |A_{n,t}|^{\xi} e^{\frac{1}{2}\xi^2 t}, \end{aligned}$$

where we have used independence of the Brownian increments and the moment generating function of the normal distribution to compute the last line. The remainder of the statement of the lemma comes from the identity $|A_{n,t}| = \frac{t^{n-1}}{(n-1)!}$ and Stirling's approximation to $n!$. $\square$

Bounds from the GUE connection

Let $\lambda_{GUE,n}$ be the top eigenvalue of an $n \times n$ GUE random matrix with entries that have variance $\sigma^2 = \frac{1}{4n}$. Then [7, Theorem 0.7] and [38] give

$$\lambda_{GUE,n} \overset{d}{=} \frac{1}{2\sqrt{n}} \max_{0=u_0<u_1<\cdots<u_{n-1}<u_n=1}\left\{\sum_{i=1}^{n} B_i(u_{i-1},u_i)\right\}.$$

The right tail rate function of $\lambda_{GUE,n}$ can be computed ([64, (1.25)], [9]) for $r > 0$ to be

$$J_{GUE}(r) = \lim_{n\to\infty} -\frac{1}{n}\log P\left(\lambda_{GUE,n} \geqslant 1+r\right) = 4\int_0^r \sqrt{x(x+2)}dx. \tag{2.4.26}$$

Lemma 2.4.22. *Suppose that $r,s,t > 0$ and $(s_n,t_n,r_n) \in \mathbb{N}\times(0,\infty)\times\mathbb{R}$ satisfy $n^{-1}(s_n,t_n,r_n) \to (s,t,r)$. If $r - s\log t - s + s\log s > 2\sqrt{ts}$, then*

$$\liminf_{n\to\infty} -\frac{1}{n}\log P\left(\log Z_{0,s_n}(0,t_n) \geqslant r_n\right) \geqslant sJ_{GUE}\left(\frac{r - s\log t - s + s\log s)}{2\sqrt{ts}} - 1\right)$$

and if $r + s\log t + s - s\log s > 2\sqrt{ts}$, *then*

$$\liminf_{n\to\infty} -\frac{1}{n}\log P\left(\log Z_{0,s_n}(0,t_n) \leqslant -r_n\right) \geqslant sJ_{GUE}\left(\frac{r + s\log t + s - s\log s}{2\sqrt{ts}} - 1\right).$$

Proof. Observe that

$$|A_{n+1,t}| = \frac{t^n}{n!} \leqslant \frac{1}{\sqrt{2\pi n}}\left(\frac{te}{n}\right)^n.$$

Using this fact and bounding $Z_{0,n}(0,t)$ as defined in (2.2.2) above with the maximum value of the Brownian increments, we obtain

$$\begin{aligned}\log Z_{0,s_n}(0,t_n) &\leqslant \log\left(\frac{1}{\sqrt{2\pi s_n}}\left(\frac{t_n e}{s_n}\right)^{s_n}\right) + \max_{0=u_0<u_1<\cdots<u_{s_n}=t_n}\left\{\sum_{i=0}^{s_n-1} B_i(u_i,u_{i+1})\right\}\\ &\overset{d}{=} \log\left(\frac{1}{\sqrt{2\pi s_n}}\left(\frac{t_n e}{s_n}\right)^{s_n}\right) + 2\sqrt{t_n s_n}\lambda_{GUE,s_n}.\end{aligned}$$

The result then follows from the inequality

$$P\left(\log Z_{0,s_n}(0,t_n) \geqslant r_n\right) \leqslant P\left(\lambda_{GUE,s_n} \geqslant \frac{r_n - s_n\log t_n - s_n + s_n\log s_n}{2\sqrt{t_n s_n}} - \frac{1}{2\sqrt{t_n s_n}}\log\left(\frac{1}{\sqrt{2\pi s_n}}\right)\right).$$

The proof of the second bound follows a similar argument: we bound the partition function below with the minimum of the Brownian increments, apply the upper bound from Stirling's approximation to $n!$, and appeal to Brownian reflection symmetry. □

Lemma 2.4.23. *Fix* $\epsilon > 0$ *and let* $s \in \mathbb{N}$ *and* $t_n = O(n^\alpha)$ *for some* $\alpha < 1$. *Then there exist* $c, C > 0$ *so that*

$$P\left(\max_{0=u_0<u_1<\cdots<u_{s-1}<u_s=t_n}\left\{\sum_{i=0}^{s-1} B_i(u_i,u_{i+1})\right\} \geqslant n\epsilon\right) \leqslant Ce^{-cn^{2-\alpha}}.$$

Proof. Large deviation estimates for largest eigenvalues give the result. For example, by [64, (2.7)], there exist $C, c > 0$ such that

$$P\left(\max_{0=u_0<u_1<\cdots<u_{s-1}<u_s=t_n}\left\{\sum_{i=0}^{s-1} B_i(u_i,u_{i+1}) \geqslant n\epsilon\right\}\right) = P\left(\lambda_{GUE,s} \geqslant \frac{n}{\sqrt{t_n}}\frac{\epsilon}{\sqrt{s}}\right) \leqslant Ce^{-cn^{2-\alpha}}.$$

□

Upper tail coarse graining estimate

Lemma 2.4.24. *Fix $a \in [0,t)$ and $\epsilon > 0$. Then for $\nu < \min(\epsilon, t-a)$*

$$P\left(\log n \int_a^{a+\nu} \frac{Z_0^\theta(nu)}{Z_0^\theta(na)} \cdot \frac{Z_{1,\lfloor ns \rfloor}(nu,nt)}{Z_{1,\lfloor ns \rfloor}(na,nt)} du \geqslant n\epsilon\right) \leqslant \exp\left\{-n\frac{1}{4}\left(\frac{\epsilon-\theta\nu}{\sqrt{\nu}}\right)^2 + o(n)\right\}.$$

Proof. We have for all $u \in (a, a+\nu)$

$$Z_{1,1}(na,nu)^{-1} Z_{1,\lfloor ns \rfloor}(nu,nt)^{-1} \geqslant Z_{1,\lfloor ns \rfloor}(na,nt)^{-1}$$

so it follows that

$$\begin{aligned} P&\left(\log n \int_a^{a+\nu} \frac{Z_0^\theta(nu)}{Z_0^\theta(na)} \frac{Z_{1,\lfloor ns \rfloor}(nu,nt)}{Z_{1,\lfloor ns \rfloor}(na,nt)} du \geqslant n\epsilon\right) \\ &\leqslant P\left(\log n \int_a^{a+\nu} \frac{Z_0^\theta(nu)}{Z_0^\theta(na)} Z_{1,1}(na,nu)^{-1} du \geqslant n\epsilon\right) \\ &= P\left(\log n \int_a^{a+\nu} e^{\theta n(u-a) - B(na,nu) - B_1(na,nu)} du \geqslant n\epsilon\right) \\ &\leqslant P\left(\max_{0 \leqslant u \leqslant 1} \{B(u) + B_1(u)\} \geqslant \sqrt{n}\left(\frac{\epsilon-\theta\nu}{\sqrt{\nu}}\right) - \frac{\log(n\nu)}{\sqrt{n\nu}}\right), \end{aligned}$$

where the last inequality comes from Brownian translation invariance, symmetry, and scaling. Recall that $B + B_1$ has the same process level distribution as $\sqrt{2}B$. The result follows from the reflection principle. $\square$

Left tail error bound

Lemma 2.4.25. *Take sequences t_n, s_n, r_n such that there exist $a, b > 0$ with $a < t_n < b$, $r_n \to r > 0$ and $s_n \in \mathbb{N}$ satisfies $s_n \log(s_n) = o(n)$. Then there exist constants $c, C > 0$ such that*

$$P\left(\log Z_{0,s_n}(0,t_n) \leqslant -nr_n\right) \leqslant Ce^{-cn^2}.$$

Proof. We have $Z_{0,s_n}(0,t_n) \geqslant \prod_{i=0}^{s_n-1} Z_{i,i+1}\left(i\frac{t_n}{s_n}, (i+1)\frac{t_n}{s_n}\right)$ where the $Z_{i,i+1}\left(i\frac{t_n}{s_n}, (i+1)\frac{t_n}{s_n}\right)$ are i.i.d.. As above in (2.4.25), there exist i.i.d. random variables $X_i \sim N\left(\log\left(\frac{t_n}{s_n}\right), \frac{2t_n}{3s_n}\right)$ with

$Z_{i,i+1}\left(i\frac{t_n}{s_n},(i+1)\frac{t_n}{s_n}\right)\geqslant X_i$. It follows that

$$P\left(\log Z_{0,s_n}(0,t_n)\leqslant -nr_n\right)\leqslant P\left(\sum_{i=0}^{s_n-1}X_i\leqslant -nr_n\right)=P\left(N\left(0,1\right)\geqslant n\frac{r_n}{\sqrt{3t_n}}+\frac{s_n}{\sqrt{3t_n}}\log\left(\frac{t_n}{s_n}\right)\right).$$

Recall that $\frac{r_n}{\sqrt{3t_n}}+\frac{s_n}{n\sqrt{3t_n}}\log\left(\frac{t_n}{s_n}\right)$ is a bounded sequence and without loss of generality is bounded away from zero. The result follows from normal tail estimates. $\square$

2.5 Large deviations for inhomogeneous exponential last passage percolation

2.5.1 Variational formulas for the Lyapunov exponents

Our purpose in this section is to prove Theorem 2.2.16.

Lemma 2.5.1. *Let $\lambda \in \mathbb{R}$. Suppose that $z > -\underline{\alpha}$ in (2.5.1), (2.5.3), and $z < \underline{\beta}$ in (2.5.2) and (2.5.4) below.*

(a) μ-a.s., for any $t > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[\exp\left(\lambda \sum_{i=1}^{\lfloor nt \rfloor} W(i,0)\right)\right] = \begin{cases} t\,\mathrm{E}\left[\log\left(\dfrac{a+z}{a+z-\lambda}\right)\right] & \textit{if } \lambda \leqslant \underline{\alpha} + z \\ \infty & \textit{otherwise.} \end{cases} \tag{2.5.1}$$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[\exp\left(\lambda \sum_{i=1}^{\lfloor nt \rfloor} W(0,i)\right)\right] = \begin{cases} t\,\mathrm{E}\left[\log\left(\dfrac{b-z}{b-z-\lambda}\right)\right] & \textit{if } \lambda \leqslant \underline{\beta} - z \\ \infty & \textit{otherwise.} \end{cases} \tag{2.5.2}$$

(b) For any $t > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}^{z}\left[\exp\left(\lambda \sum_{i=1}^{\lfloor nt \rfloor} W(i,0)\right)\right] = \begin{cases} t\log\mathrm{E}\left[\dfrac{a+z}{a+z-\lambda}\right] & \textit{if } \lambda \leqslant \underline{\alpha} + z \\ \infty & \textit{otherwise.} \end{cases} \tag{2.5.3}$$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}^{z}\left[\exp\left(\lambda \sum_{i=1}^{\lfloor nt \rfloor} W(0,i)\right)\right] = \begin{cases} t\log\mathrm{E}\left[\dfrac{b-z}{b-z-\lambda}\right] & \textit{if } \lambda \leqslant \underline{\beta} - z \\ \infty & \textit{otherwise.} \end{cases} \tag{2.5.4}$$

Proof. Using (2.2.25), we compute

$$\mathbf{E}^z_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum_{i=1}^{\lfloor nt\rfloor} W(i,0)}\right] = \begin{cases} \prod_{i=1}^{\lfloor nt\rfloor} \dfrac{a_i+z}{a_i+z-\lambda} & \text{if } \lambda < \min\limits_{1\leqslant i\leqslant \lfloor nt\rfloor} a_i + z \\ \infty & \text{otherwise.} \end{cases} \tag{2.5.5}$$

If $\lambda < \underline{a} + z$ then the first equality in (2.5.5) holds for all $n \in \mathbb{N}$ μ-a.s and we have

$$\mathrm{E}\left|\log\left(\frac{a+z}{a+z-\lambda}\right)\right| < \infty. \tag{2.5.6}$$

Hence, by the ergodicity of $\mathbf{a}$,

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbf{E}^z_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum_{i=1}^{\lfloor nt\rfloor} W(i,0)}\right] = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor}\log\left(\frac{a_i+z}{a_i+z-\lambda}\right) = t\,\mathrm{E}\log\left(\frac{a+z}{a+z-\lambda}\right) \quad \mu\text{-a.s.} \tag{2.5.7}$$

Moreover, it follows from (2.5.5) that

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}^z\left[e^{\lambda\sum_{i=1}^{\lfloor nt\rfloor} W(i,0)}\right] = \lim_{n\to\infty}\frac{\lfloor nt\rfloor}{n}\log\mathrm{E}\left[\frac{a+z}{a+z-\lambda}\right] \to t\log\mathrm{E}\left[\frac{a+z}{a+z-\lambda}\right]. \tag{2.5.8}$$

Next, consider the case $\lambda = \underline{a} + z$. If (2.5.6) is in force, then both (2.5.7) and (2.5.8) still hold. Suppose now that (2.5.6) fails. By monotonicity,

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbf{E}^z_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum_{i=1}^{\lfloor nt\rfloor} W(i,0)}\right] \geqslant t\,\mathrm{E}\log\left(\frac{a+z}{a+z-\lambda'}\right) \quad \mu\text{-a.s.}$$
$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}^z\left[e^{\lambda\sum_{i=1}^{\lfloor nt\rfloor} W(i,0)}\right] \geqslant t\log\mathrm{E}\left[\frac{a+z}{a+z-\lambda'}\right]$$

for any $\lambda' < \lambda$. Letting $\lambda' \uparrow \lambda$ and monotone convergence yield

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbf{E}^z_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum_{i=1}^{\lfloor nt\rfloor} W(i,0)}\right] = \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}^z\left[e^{\lambda\sum_{i=1}^{\lfloor nt\rfloor} W(i,0)}\right] = \infty. \tag{2.5.9}$$

Finally, consider the case $\lambda > \underline{a} + z$. Then, by the ergodicity of $\mathbf{a}$, there exists $i \in \mathbb{N}$ such that $\lambda \geqslant a_i + z$ and the second equality in (2.5.5) holds for large enough $n \in \mathbb{N}$ μ-a.s. Hence, (2.5.9).

We have verified (2.5.1) and (2.5.3). The proofs of (2.5.2) and (2.5.4) are similar. $\square$

Recall the basic properties of the Lyapunov exponents stated in Proposition 2.5.15. For $s, t > 0$ and $\lambda \in \mathbb{R}$, define $\mathbf{L}_{s,0}(\lambda) = \lim_{t\downarrow 0} \mathbf{L}_{s,t}(\lambda)$ and $\mathbf{L}_{0,t}(\lambda) = \lim_{s\downarrow 0} \mathbf{L}_{s,t}(\lambda)$, where the limits exist by monotonicity. Define $\mathbb{L}_{s,0}(\lambda)$ and $\mathbb{L}_{0,t}(\lambda)$ similarly. Also, for $k, l \in \mathbb{Z}_+$, let $\theta_{k,l}$ denote the shift given by $\omega(i,j) \mapsto \omega(i+k, j+l)$ for $i, j \in \mathbb{N}$ and $\omega \in \mathbb{R}^{\mathbb{N}^2}$. We next obtain a variational formula involving the Lyapunov exponents.

Lemma 2.5.2. *Let $z \in (-\underline{\alpha}, \underline{\beta})$ and $\lambda \in (0, \underline{\beta} - z]$. Then*

$$\begin{aligned}
&\mathrm{E}\log\left(\frac{a+z+\lambda}{a+z}\right) + \mathrm{E}\log\left(\frac{b-z}{b-z-\lambda}\right) \qquad (2.5.10)\\
&= \sup_{0\leqslant t\leqslant 1}\left\{\max\left\{\mathbf{L}_{t,1}(\lambda) + (1-t)\,\mathrm{E}\log\left(\frac{a+z+\lambda}{a+z}\right), \mathbf{L}_{1,t}(\lambda) + (1-t)\,\mathrm{E}\log\left(\frac{b-z}{b-z-\lambda}\right)\right\}\right\}.
\end{aligned}$$

Also,

$$\begin{aligned}
&\log\mathrm{E}\left[\frac{a+z+\lambda}{a+z}\right] + \log\mathrm{E}\left[\frac{b-z}{b-z-\lambda}\right] \qquad (2.5.11)\\
&= \sup_{0\leqslant t\leqslant 1}\left\{\max\left\{\mathbb{L}_{t,1}(\lambda) + (1-t)\log\mathrm{E}\left[\frac{a+z+\lambda}{a+z}\right], \mathbb{L}_{1,t}(\lambda) + (1-t)\log\mathrm{E}\left[\frac{b-z}{b-z-\lambda}\right]\right\}\right\}.
\end{aligned}$$

Proof of (2.5.10). We may assume that the left-hand side of (2.5.10) is finite. (This assumption fails only when $\lambda = \underline{\beta} - z$ and $\mathrm{E}\log(b - \underline{\beta}) = -\infty$ in which case (2.5.10) clearly holds).

It follows from (2.2.13) and (2.2.24) that

$$\widehat{G}(n,n) = \max_{1\leqslant k\leqslant n}\{\max\{G(n-k+1,n)\circ\theta_{k-1,0} + \widehat{G}(k,0), G(n,n-k+1)\circ\theta_{0,k-1} + \widehat{G}(0,k)\}\},$$

which leads to

$$\begin{aligned}
\sum_{1\leqslant j\leqslant n} J(n,j) = \max_{1\leqslant k\leqslant n}\{\max\{&G(n-k+1,n)\circ\theta_{k-1,0} - \sum_{k<i\leqslant n} W(i,0),\\
&G(n,n-k+1)\circ\theta_{0,k-1} - \sum_{1\leqslant i\leqslant n} W(i,0) + \sum_{1\leqslant j\leqslant k} W(0,j)\}\}.
\end{aligned} \qquad (2.5.12)$$

Also, note the identity

$$\frac{1}{\mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{-\lambda W(i,0)}\right]} = \frac{a_i+z+\lambda}{a_i+z} = \mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda W(i,0)}\right] \text{ for } \lambda > 0 \text{ and } z > -\underline{\alpha}. \qquad (2.5.13)$$

Using the independence of weights under $\mathbf{P}^{z}_{\mathbf{a},\mathbf{b}}$, Proposition 2.2.31, (2.5.12) and (2.5.13), we obtain

$$\begin{aligned}
&\mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant i\leqslant n}W(i,0)}\right]\cdot\mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant j\leqslant n}W(0,j)}\right]\\
&\geqslant\max\left\{\mathbf{E}_{\tau_{k-1}(\mathbf{a}),\mathbf{b}}\left[e^{\lambda G(n-k+1,n)}\right]\cdot\mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant i\leqslant k}W(i,0)}\right],\right.\\
&\qquad\left.\mathbf{E}_{\mathbf{a},\tau_{k-1}(\mathbf{b})}\left[e^{\lambda G(n,n-k+1)}\right]\cdot\mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant j\leqslant k}W(0,j)}\right]\right\}.
\end{aligned}\tag{2.5.14}$$

Set $k=\lceil n(1-t)\rceil+1$ for some $t\in(0,1)$, apply logarithms to both sides and divide through by n in (2.5.14). It follows from Proposition 2.5.15 that

$$\frac{1}{n}\log\mathbf{E}_{\tau_{k-1}(\mathbf{a}),\mathbf{b}}\left[e^{\lambda G(n-k+1,n)}\right]\to\mathbf{L}_{t,1}(\lambda),\qquad\frac{1}{n}\log\mathbf{E}_{\mathbf{a},\tau_{k-1}(\mathbf{b})}\left[e^{\lambda G(n,n-k+1)}\right]\to\mathbf{L}_{1,t}(\lambda)$$

as $n\to\infty$ along suitable subsequences because $(\mathbf{a},\mathbf{b})$ is stationary and $\mathbf{L}$ is deterministic. Hence, also using Lemma 2.5.1, we obtain

$$\begin{aligned}
&\mathrm{E}\log\left(\frac{a+z+\lambda}{a+z}\right)+\mathrm{E}\log\left(\frac{b-z}{b-z-\lambda}\right)\\
&\geqslant\max\left\{\mathbf{L}_{t,1}(\lambda)+(1-t)\,\mathrm{E}\log\left(\frac{a+z+\lambda}{a+z}\right),\mathbf{L}_{1,t}(\lambda)+(1-t)\,\mathrm{E}\log\left(\frac{b-z}{b-z-\lambda}\right)\right\}.
\end{aligned}\tag{2.5.15}$$

In particular, $\mathbf{L}$ is finite. By continuity, (2.5.15) holds with $t=0$ and $t=1$ as well.

For the opposite inequality, introduce $L\in\mathbb{N}$ and let $n>L$ such that $\lceil(l+1)n/L\rceil>\lceil ln/L\rceil$ for $0\leqslant l<L$. Then, by (2.5.12) and nonnegativity of the weights,

$$\begin{aligned}
\sum_{1\leqslant j\leqslant n}J(n,j)\leqslant\max_{1\leqslant l<L}\{\max\{&G(\lfloor(L-l)n/L\rfloor,n)\circ\theta_{\lceil ln/L\rceil,0}-\sum_{\lceil(l+1)n/L\rceil<i\leqslant n}W(i,0),\\
&G(n,\lfloor(L-l)n/L\rfloor)\circ\theta_{0,\lceil ln/L\rceil}-\sum_{1\leqslant i\leqslant n}W(i,0)+\sum_{1\leqslant j\leqslant\lceil(l+1)n/L\rceil}W(0,j)\},
\end{aligned}$$

which implies that

$$\begin{aligned} &\mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant i\leqslant n}W(i,0)}\right]\cdot\mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant j\leqslant n}W(0,j)}\right] \\ &\leqslant\sum_{0\leqslant l<L}\mathbf{E}_{\tau_{\lceil ln/L\rceil}(\mathbf{a}),\mathbf{b}}\left[e^{\lambda G(\lfloor(L-l)n/L\rfloor,n)}\right]\cdot\mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{i=1}^{\lceil(l+1)n/L\rceil}W(i,0)}\right] \\ &\qquad+\mathbf{E}_{\mathbf{a},\tau_{\lceil ln/L\rceil}(\mathbf{b})}\left[e^{\lambda G(n,\lfloor(L-l)n/L\rfloor)}\right]\cdot\mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{j=1}^{\lceil(l+1)n/L\rceil}W(0,j)}\right]. \end{aligned} \tag{2.5.16}$$

Taking logarithms leads to

$$\begin{aligned} &\log\mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant i\leqslant n}W(i,0)}\right]+\log\mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant j\leqslant n}W(0,j)}\right] \\ &\leqslant\max_{0\leqslant l<L}\max\Bigg\{\log\mathbf{E}_{\tau_{\lceil ln/L\rceil}(\mathbf{a}),\mathbf{b}}\left[e^{\lambda G(\lfloor(L-l)n/L\rfloor,n)}\right]+\log\mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{i=1}^{\lceil(l+1)n/L\rceil}W(i,0)}\right], \\ &\qquad\log\mathbf{E}_{\mathbf{a},\tau_{\lceil ln/L\rceil}(\mathbf{b})}\left[e^{\lambda G(n,\lfloor(L-l)n/L\rfloor)}\right]+\log\mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{j=1}^{\lceil(l+1)n/L\rceil}W(0,j)}\right]\Bigg\}+\log(2L). \end{aligned}$$

Dividing through by n and letting $n\to\infty$ along a suitable subsequential limit yield

$$\begin{aligned} &\mathrm{E}\log\left(\frac{a+z+\lambda}{a+z}\right)+\mathrm{E}\log\left(\frac{b-z}{b-z-\lambda}\right) \\ &\leqslant\max_{0\leqslant l<L}\max\left\{\mathbf{L}_{1-l/L,1}(\lambda)+\frac{l+1}{L}\mathrm{E}\log\left(\frac{a+z+\lambda}{a+z}\right),\mathbf{L}_{1,1-l/L}(\lambda)+\frac{l+1}{L}\mathrm{E}\log\left(\frac{b-z}{b-z-\lambda}\right)\right\} \\ &\leqslant\sup_{0\leqslant t\leqslant 1}\max\left\{\mathbf{L}_{t,1}(\lambda)+(1-t)\mathrm{E}\log\left(\frac{a+z+\lambda}{a+z}\right),\mathbf{L}_{1,t}(\lambda)+(1-t)\mathrm{E}\log\left(\frac{b-z}{b-z-\lambda}\right)\right\} \\ &+\frac{1}{L}\left(\mathrm{E}\log\left(\frac{a+z+\lambda}{a+z}\right)+\mathrm{E}\log\left(\frac{b-z}{b-z-\lambda}\right)\right) \end{aligned}$$

Letting $L\to\infty$ completes the proof. □

Proof of (2.5.11). Some details are skipped. We may assume that the left-hand side of (2.5.11) is finite.

Using independence, we can rewrite (2.5.14) as

$$\begin{aligned} \mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{k<i\leqslant n}W(i,0)}\right]\cdot\mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant j\leqslant n}W(0,j)}\right]&\geqslant\mathbf{E}_{\tau_{k-1}(\mathbf{a}),\mathbf{b}}\left[e^{\lambda G(n-k+1,n)}\right] \\ \mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant i\leqslant n}W(i,0)}\right]\cdot\mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{k<j\leqslant n}W(0,j)}\right]&\geqslant\mathbf{E}_{\mathbf{a},\tau_{k-1}(\mathbf{b})}\left[e^{\lambda G(n,n-k+1)}\right] \end{aligned} \tag{2.5.17}$$

The factors on the right-hand side are independent. Applying E yields

$$\begin{aligned}
&\mathbb{E}^{z+\lambda}\left[e^{\lambda\sum\limits_{1\leqslant i\leqslant n}W(i,0)}\right]\cdot\mathbb{E}^{z}\left[e^{\lambda\sum\limits_{1\leqslant j\leqslant n}W(0,j)}\right] \qquad (2.5.18)\\
&\geqslant \max\left\{\mathbb{E}\left[e^{\lambda G(n-k+1,n)}\right]\cdot\mathbb{E}^{z+\lambda}\left[e^{\lambda\sum\limits_{1\leqslant i\leqslant k}W(i,0)}\right],\mathbb{E}\left[e^{\lambda G(n,n-k+1)}\right]\cdot\mathbb{E}^{z}\left[e^{\lambda\sum\limits_{1\leqslant j\leqslant k}W(0,j)}\right]\right\},
\end{aligned}$$

where we rearranged terms using that $\{W(i,0) : i \in \mathbb{N}\}$ and $\{W(0,j) : j \in \mathbb{N}\}$ are both i.i.d. under $\mathbb{P}^{z+\lambda}$ and $\mathbb{P}^{z}$. Then, (2.5.18) leads to $\geqslant$ half of (2.5.11) via Proposition 2.5.15 and Lemma 2.5.1.

For the $\leqslant$ half of (2.5.11), suppose that $\lambda < \underline{\beta} - z$ for the moment. Note the inequalities

$$\mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}[e^{\lambda W(i,0)}] = \frac{a_i+z+\lambda}{a_i+z} \leqslant \frac{\underline{\alpha}+z+\lambda}{\underline{\alpha}+z}, \qquad \mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}[e^{\lambda W(0,j)}] = \frac{b_j-z}{b_j-z-\lambda} \leqslant \frac{\underline{\beta}-z}{\underline{\beta}-z-\lambda}.$$

It follows from these and (2.5.16) that

$$\begin{aligned}
&\mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant i\leqslant n}W(i,0)}\right]\cdot\mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{1\leqslant j\leqslant n}W(0,j)}\right]\\
&\leqslant \sum_{0\leqslant l<L}\left(\frac{\underline{\alpha}+z+\lambda}{\underline{\alpha}+z}\right)^{n/L+1}\mathbf{E}_{\tau_{\lceil ln/L\rceil}(\mathbf{a}),\mathbf{b}}\left[e^{\lambda G(\lfloor (L-l)n/L\rfloor,n)}\right]\cdot\mathbf{E}^{z+\lambda}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{i=1}^{\lceil ln/L\rceil}W(i,0)}\right] \qquad (2.5.19)\\
&\qquad+\left(\frac{\underline{\beta}-z}{\underline{\beta}-z-\lambda}\right)^{n/L+1}\mathbf{E}_{\mathbf{a},\tau_{\lceil ln/L\rceil}(\mathbf{b})}\left[e^{\lambda G(n,\lfloor (L-l)n/L\rfloor)}\right]\cdot\mathbf{E}^{z}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda\sum\limits_{j=1}^{\lceil ln/L\rceil}W(0,j)}\right].
\end{aligned}$$

The point of (2.5.19) is that the terms on the right-hand side are products of independent factors, which is not the case in (2.5.16). Applying $\log \mathrm{E}$, we obtain

$$\begin{aligned}
&\log\mathbb{E}^{z+\lambda}\left[e^{\lambda\sum\limits_{1\leqslant i\leqslant n}W(i,0)}\right]+\log\mathbb{E}^{z}\left[e^{\lambda\sum\limits_{1\leqslant j\leqslant n}W(0,j)}\right]\\
&\leqslant \max_{0\leqslant l<L}\max\left\{(n/L+1)\log\left(\frac{\underline{\alpha}+z+\lambda}{\underline{\alpha}+z}\right)+\log\mathbb{E}\left[e^{\lambda G(\lfloor (L-l)n/L\rfloor,n)}\right]-\log\mathbb{E}^{z+\lambda}\left[e^{\lambda\sum\limits_{i=1}^{\lceil ln/L\rceil}W(i,0)}\right],\right.\\
&\qquad\left.(n/L+1)\log\left(\frac{\underline{\beta}-z}{\underline{\beta}-z-\lambda}\right)+\log\mathbb{E}\left[e^{\lambda G(n,\lfloor (L-l)n/L\rfloor)}\right]+\log\mathbb{E}^{z}\left[e^{\lambda\sum\limits_{j=1}^{\lceil ln/L\rceil}W(0,j)}\right]\right\}\\
&\qquad+\log(2L).
\end{aligned}$$

Divide through by n and let $n \to \infty$. If we then send $L \to \infty$, the result is

$$\log \mathrm{E}\left[\frac{a+z+\lambda}{a+z}\right] + \log \mathrm{E}\left[\frac{b-z}{b-z-\lambda}\right] \leqslant \sup_{0\leqslant t\leqslant 1}\left\{\max\left\{\mathbb{L}_{t,1}(\lambda) + (1-t)\log \mathrm{E}\left[\frac{a+z+\lambda}{a+z}\right],\right.\right.$$
$$\left.\left.\mathbb{L}_{t,1}(\lambda) + (1-t)\log \mathrm{E}\left[\frac{b-z}{b-z-\lambda}\right]\right\}\right\}.$$

for all $\lambda < \underline{\beta} - z$. The case $\lambda = \underline{\beta} - z$ also follows because the right-hand side is nondecreasing in λ and the left-hand side, due to monotone convergence, is continuous in λ on $(0, \underline{\beta} - z]$. $\square$

Lemma 2.5.3. *For $\lambda > 0$,*

$$\mathbf{L}_{1,0}(\lambda) = \mathrm{E}\log\left(\frac{a+\underline{\beta}}{a+\underline{\beta}-\lambda}\right) \quad \mathbf{L}_{0,1}(\lambda) = \mathrm{E}\log\left(\frac{b+\underline{\alpha}}{b+\underline{\alpha}-\lambda}\right) \qquad \text{if } \lambda \leqslant \underline{\alpha} + \underline{\beta} \tag{2.5.20}$$

$$\mathbf{L}_{1,0}(\lambda) = \mathbf{L}_{0,1}(\lambda) = \infty \qquad \text{otherwise.} \tag{2.5.21}$$

$$\mathbb{L}_{1,0}(\lambda) = \log \mathrm{E}\left[\frac{a+\underline{\beta}}{a+\underline{\beta}-\lambda}\right] \quad \mathbb{L}_{0,1}(\lambda) = \log \mathrm{E}\left[\frac{b+\underline{\alpha}}{b+\underline{\alpha}-\lambda}\right] \qquad \text{if } \lambda \leqslant \underline{\alpha} + \underline{\beta} \tag{2.5.22}$$

$$\mathbb{L}_{1,0}(\lambda) = \mathbb{L}_{0,1}(\lambda) = \infty \qquad \text{otherwise.} \tag{2.5.23}$$

Proof. Let $\epsilon > 0$. On the event $b_1 \leqslant \underline{\beta} + \epsilon$, which has positive μ-probability, we have for $n \geqslant 1/\epsilon$

$$\begin{aligned}
\frac{1}{n}\log \mathbf{E}_{\mathbf{a},\mathbf{b}}[e^{\lambda G(n,\lfloor n\epsilon\rfloor)}] &\geqslant \frac{1}{n}\log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda \sum_{1\leqslant i\leqslant n} W(i,1)}\right] \\
&= \begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} \dfrac{a_i+b_1}{a_i+b_1-\lambda} & \text{if } \lambda < \min\limits_{1\leqslant i\leqslant n} a_i + b_1 \\ \infty & \text{otherwise} \end{cases} \\
&\geqslant \begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} \dfrac{a_i+\underline{\beta}+\epsilon}{a_i+\underline{\beta}+\epsilon-\lambda} & \text{if } \lambda < \min\limits_{1\leqslant i\leqslant n} a_i + \underline{\beta}+\epsilon \\ \infty & \text{otherwise} \end{cases} \\
&= \frac{1}{n}\log \mathbf{E}^{\underline{\beta}+\epsilon}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda \sum_{1\leqslant i\leqslant n} W(i,0)}\right].
\end{aligned}$$

Then, by Lemma 2.5.1,

$$\mathbf{L}_{1,\epsilon}(\lambda) \geqslant \begin{cases} \mathrm{E}\left[\log\left(\dfrac{a+\underline{\beta}+\epsilon}{a+\underline{\beta}+\epsilon-\lambda}\right)\right] & \text{if } \lambda \leqslant \underline{\alpha}+\underline{\beta}+\epsilon \\ \infty & \text{otherwise.} \end{cases}.$$

By monotone convergence, letting $\epsilon \downarrow 0$ yields

$$\mathbf{L}_{1,0}(\lambda) \geqslant \begin{cases} \mathrm{E}\left[\log\left(\dfrac{a+\underline{\beta}}{a+\underline{\beta}-\lambda}\right)\right] & \text{if } \lambda \leqslant \underline{\alpha}+\underline{\beta} \\ \infty & \text{otherwise.} \end{cases}.$$

To complete the proof of (2.5.21), we need

$$\mathbf{L}_{1,0}(\lambda) \leqslant \mathrm{E}\left[\log\left(\frac{a+\underline{\beta}}{a+\underline{\beta}-\lambda}\right)\right] \tag{2.5.24}$$

for $\lambda \in (0, \underline{\alpha}+\underline{\beta}]$. When $\lambda = \underline{\alpha}+\underline{\beta}$, we may assume that the right-hand side is finite. Then, $a_i > \underline{\alpha}$ for $i \in \mathbb{N}$ a.s. and the argument in the paragraph of inequality (2.5.14) goes through with $z = -\underline{\alpha}$ as well. Hence,

$$\mathrm{E}\left[\log\left(\frac{a+z+\lambda}{a+z}\right)\right] + \mathrm{E}\left[\log\left(\frac{b-z}{b-z-\lambda}\right)\right] \geqslant \mathbf{L}_{1,t}(\lambda) + (1-t)\,\mathrm{E}\left[\log\left(\frac{b-z}{b-z-\lambda}\right)\right] \tag{2.5.25}$$

for $t \in [0,1]$, $z \in [-\underline{\alpha}, \underline{\beta})$ and $\lambda \in (0, \underline{\beta}-z]$, which simplifies to

$$\mathrm{E}\left[\log\left(\frac{a+z+\lambda}{a+z}\right)\right] + t\,\mathrm{E}\left[\log\left(\frac{b-z}{b-z-\lambda}\right)\right] \geqslant \mathbf{L}_{1,0}(\lambda). \tag{2.5.26}$$

Setting $t = 0$ and $z = \underline{\beta}-\lambda$ in (2.5.26) gives (2.5.24). The remaining cases are treated similarly. □

Corollary 2.5.4. *For $s, t > 0$,*

$$\mathbf{L}_{s,t}(\underline{\alpha}+\underline{\beta}) = s\,\mathrm{E}\log\left(\frac{a+\underline{\beta}}{a-\underline{\alpha}}\right) + t\,\mathrm{E}\log\left(\frac{b+\underline{\alpha}}{b-\underline{\beta}}\right).$$

$$\mathbb{L}_{s,t}(\underline{\alpha}+\underline{\beta}) = s\log\mathrm{E}\left[\frac{a+\underline{\beta}}{a-\underline{\alpha}}\right] + t\log\mathrm{E}\left[\frac{b+\underline{\alpha}}{b-\underline{\beta}}\right].$$

Proof. By concavity and homogeneity,

$$\mathbf{L}_{s,t}(\underline{\alpha}+\underline{\beta}) \geqslant s\,\mathbf{L}_{1,0}(\underline{\alpha}+\underline{\beta}) + t\,\mathbf{L}_{0,1}(\underline{\alpha}+\underline{\beta}) = s\,\mathrm{E}\left[\log\left(\frac{a+\underline{\beta}}{a-\underline{\alpha}}\right)\right] + t\,\mathrm{E}\left[\log\left(\frac{b+\underline{\alpha}}{b-\underline{\beta}}\right)\right]. \tag{2.5.27}$$

When the right-hand side is finite, the opposite inequality comes from (2.5.25). $\mathbb{L}_{s,t}(\underline{\alpha}+\underline{\beta})$ is computed similarly. ☐

Proof of Theorem 2.2.16. It follows from Lemma 2.5.3 that $\mathbf{L}_{s,t}(\lambda) = \infty$ for $\lambda > \underline{\alpha}+\underline{\beta}$. Fix $\lambda \in (0, \underline{\alpha}+\underline{\beta})$ and define

$$A(z) = \mathrm{E}\left[\log\left(\frac{a+z+\lambda}{a+z}\right)\right] \text{ for } z > -\underline{\alpha}, \qquad B(z) = \mathrm{E}\left[\log\left(\frac{b-z}{b-z-\lambda}\right)\right] \text{ for } z < \underline{\beta}-\lambda.$$

Lemma 2.5.2 states that

$$A(z)+B(z) = \sup_{0\leqslant t\leqslant 1}\{\max\{\mathbf{L}_{t,1}(\lambda)+(1-t)A(z), \mathbf{L}_{1,t}(\lambda)+(1-t)B(z)\}\} \quad \text{for } z \in (-\underline{\alpha}, \underline{\beta}-\lambda).$$

Note that A and B are continuous, A is decreasing and B is increasing. Moreover, by Lemma 2.5.3, $A(\underline{\beta}-\lambda) = \mathbf{L}_{1,0}(\lambda)$ and $B(-\underline{\alpha}) = \mathbf{L}_{0,1}(\lambda)$. Also, $\mathbf{L}_{s,t}(\lambda)$ is finite and, by Proposition 2.5.15, is nondecreasing, homogeneous, concave and continuous. Thus, the setting is as in [31, Section 5] and the arguments there show that $\mathbf{L}_{s,t}(\lambda) = \inf_{-\underline{\alpha}<z<\underline{\beta}-\lambda}\{sA(z)+tB(z)\}$. The endpoints can be included in the infimum, by monotone convergence. The proof of (2.2.17) is similar. ☐

2.5.2 Extremizers of the variational problems

In this section, we derive some regularity properties of $\mathbf{L}, \mathbb{L}, \mathbf{J}$ and $\mathbb{J}$ by studying the extremizers of their variational representations. The next two lemmas describe the minimizers of (2.2.16) and (2.2.17). See Figure 10 for an illustration.

Lemma 2.5.5. *Fix $s,t>0$ and define $F = F(z,\lambda)$ for $0<\lambda<\underline{\alpha}+\underline{\beta}$ and $-\underline{\alpha} \leqslant z \leqslant \underline{\beta}-\lambda$ by*

$$F(z,\lambda) = s\,\mathrm{E}\log\left(\frac{a+z+\lambda}{a+z}\right) + t\,\mathrm{E}\log\left(\frac{b-z}{b-z-\lambda}\right). \tag{2.5.28}$$

For each $\lambda \in (0, \underline{\alpha} + \underline{\beta})$*, there exists a unique* $\mathrm{z}_\star = \mathrm{z}_\star(\lambda) \in [-\underline{\alpha}, \underline{\beta} - \lambda]$ *such that* $\mathbf{L}_{s,t}(\lambda) = F(\mathrm{z}_\star, \lambda)$*. We have* $\mathrm{z}_\star = -\underline{\alpha}$ *if and only if*

$$-s\,\mathrm{E}\left[\frac{1}{(a-\underline{\alpha}+\lambda)(a-\underline{\alpha})}\right] + t\,\mathrm{E}\left[\frac{1}{(b+\underline{\alpha}-\lambda)(b+\underline{\alpha})}\right] \geqslant 0, \tag{2.5.29}$$

and $\mathrm{z}_\star = \underline{\beta} - \lambda$ *if and only if*

$$-s\,\mathrm{E}\left[\frac{1}{(a+\underline{\beta})(a+\underline{\beta}-\lambda)}\right] + t\,\mathrm{E}\left[\frac{1}{(b-\underline{\beta})(b-\underline{\beta}+\lambda)}\right] \leqslant 0. \tag{2.5.30}$$

Define $\lambda_1 = \inf\{\lambda \in (0, \underline{\alpha}+\underline{\beta}) :$ *(2.5.29) holds.*$\} \wedge (\underline{\alpha}+\underline{\beta})$ *and* $\lambda_2 = \inf\{\lambda \in (0, \underline{\alpha}+\underline{\beta}) :$ *(2.5.30) holds.*$\} \wedge (\underline{\alpha}+\underline{\beta})$*. Then* $\mathrm{z}_\star = -\underline{\alpha}$ *if and only if* $\lambda \geqslant \lambda_1$*, and* $\mathrm{z}_\star = \underline{\beta} - \lambda$ *if and only if* $\lambda \geqslant \lambda_2$*. For* $0 < \lambda < \lambda_0 = \lambda_1 \wedge \lambda_2$*, we have* $\partial_z F(\mathrm{z}_\star, \lambda) = 0$*. Moreover,* $\mathrm{z}_\star$ *is continuous on* $(0, \underline{\alpha}+\underline{\beta})$ *and continuously differentiable on* $(0, \underline{\alpha}+\underline{\beta}) \setminus \{\lambda_0\}$*. We have* $-1 < \mathrm{z}_\star{}' < 0$ *for* $0 < \lambda < \lambda_0$*,* $\lim_{\lambda \downarrow 0} \mathrm{z}_\star = \zeta(s,t)$ *and* $\lim_{\lambda \uparrow \underline{\alpha}+\underline{\beta}} \mathrm{z}_\star = -\underline{\alpha}$*.*

Lemma 2.5.6. *Lemma 2.5.5 holds verbatim if* $\mathbf{L}_{s,t}$*, (2.5.28), (2.5.29) and (2.5.30) are replaced with* $\mathbb{L}_{s,t}$*,*

$$F(z,\lambda) = s\log \mathrm{E}\left[\frac{a+z+\lambda}{a+z}\right] + t\log \mathrm{E}\left[\frac{b-z}{b-z-\lambda}\right] \tag{2.5.31}$$

$$-s\frac{\mathrm{E}\left[\dfrac{1}{(a-\underline{\alpha})^2}\right]}{\mathrm{E}\left[\dfrac{a-\underline{\alpha}+\lambda}{a-\underline{\alpha}}\right]} + t\frac{\mathrm{E}\left[\dfrac{1}{(b+\underline{\alpha}-\lambda)^2}\right]}{\mathrm{E}\left[\dfrac{b+\underline{\alpha}}{b+\underline{\alpha}-\lambda}\right]} \geqslant 0 \tag{2.5.32}$$

$$-s\frac{\mathrm{E}\left[\dfrac{1}{(a+\underline{\beta}-\lambda)^2}\right]}{\mathrm{E}\left[\dfrac{a+\underline{\beta}}{a+\underline{\beta}-\lambda}\right]} + t\frac{\mathrm{E}\left[\dfrac{1}{(b-\underline{\beta})^2}\right]}{\mathrm{E}\left[\dfrac{b-\underline{\beta}+\lambda}{b-\underline{\beta}}\right]} \leqslant 0, \tag{2.5.33}$$

respectively. Here, the left-hand sides of (2.5.32) and (2.5.33) are interpreted as $-\infty$ *and* ∞ *when* $\mathrm{E}[(a-\underline{\alpha})^{-1}] = \infty$ *and* $\mathrm{E}[(b-\underline{\beta})^{-1}] = \infty$*, respectively.*

Proof of Lemma 2.5.5. Since $\partial_z^2 F > 0$, the existence and the uniqueness of $\mathrm{z}_\star$ follows. Also, $\mathrm{z}_\star = -\underline{\alpha}$ if and only if $\partial_z F(-\underline{\alpha}, \lambda) \geqslant 0$, which is (2.5.29). We note that $\partial_z F(-\underline{\alpha}, \lambda) = -\infty$

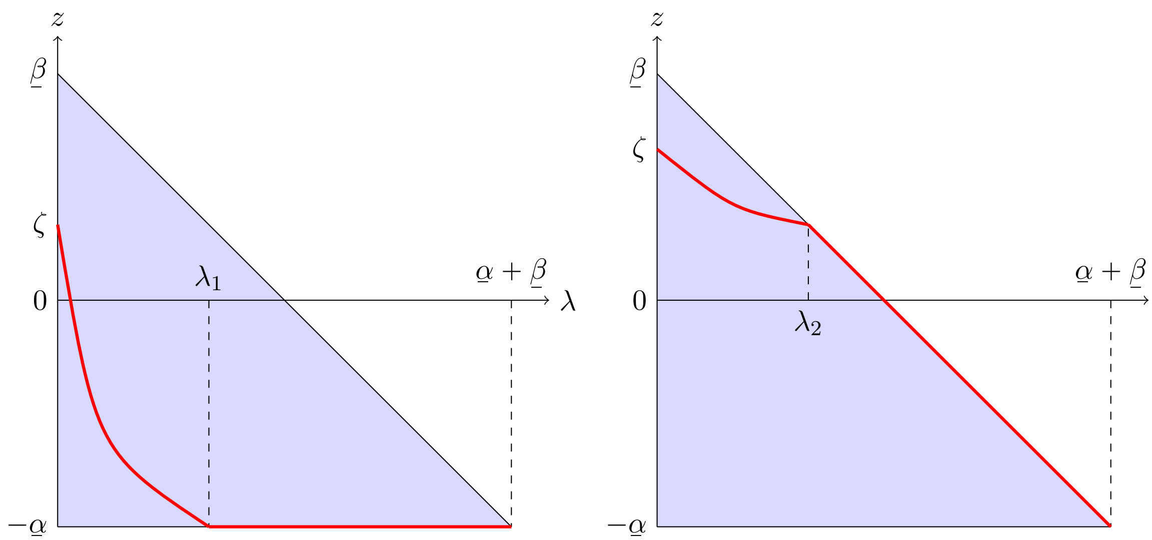


Figure 10: Sketches of the graph of the minimizers in (2.2.16) and (2.2.17) assuming (2.5.29) and (2.5.32), respectively (left) and assuming (2.5.30) and (2.5.33), respectively (right).

if $\mathrm{E}[(a-\underline{\alpha})^{-1}] = \infty$ and, otherwise, $\lambda^{-1}\partial_z F(-\underline{\alpha},\lambda)$ is a continuous, increasing function of $\lambda \in (0, \underline{\alpha}+\underline{\beta})$. Therefore, $\mathrm{z}_\star = -\underline{\alpha}$ if and only if $\lambda \geqslant \lambda_1$. We similarly observe (2.5.30) and the equivalence of $\mathrm{z}_\star = \underline{\beta} - \lambda$ and $\lambda \geqslant \lambda_2$. (Because $\partial_z F$ is increasing in z, we cannot have λ_1 and λ_2 both less than $\underline{\alpha}+\underline{\beta}$).

When $\lambda < \lambda_0$, the minimizer is the unique $\mathrm{z}_\star \in (-\underline{\alpha}, \underline{\beta}-\lambda)$ satisfying

$$\partial_z F(\mathrm{z}_\star, \lambda) = 0. \tag{2.5.34}$$

By the implicit function theorem, $\mathrm{z}_\star$ is continuously differentiable for $0 < \lambda < \lambda_0$ with derivative

$$\mathrm{z}_\star{}'(\lambda) = -\frac{\partial_\lambda \partial_z F(\mathrm{z}_\star, \lambda)}{\partial_z^2 F(\mathrm{z}_\star, \lambda)}. \tag{2.5.35}$$

Observing that

$$\partial_\lambda \partial_z F(\mathrm{z}_\star,\lambda) > -s\,\mathrm{E}\left[\frac{1}{(a+\mathrm{z}_\star)(a+\mathrm{z}_\star+\lambda)}\right] + t\,\mathrm{E}\left[\frac{1}{(b-\mathrm{z}_\star)(b-\mathrm{z}_\star-\lambda)}\right] = \lambda^{-1}\partial_z F(\mathrm{z}_\star,\lambda) = 0$$

$$\partial_z^2 F(\mathrm{z}_\star,\lambda) - \partial_\lambda \partial_z F(\mathrm{z}_\star,\lambda) = s\,\mathrm{E}\left[\frac{1}{(a+\mathrm{z}_\star)^2}\right] - t\,\mathrm{E}\left[\frac{1}{(b-\mathrm{z}_\star)^2}\right]$$

$$> s\,\mathrm{E}\left[\frac{1}{(a+\mathrm{z}_\star)(a+\mathrm{z}_\star+\lambda)}\right] - t\,\mathrm{E}\left[\frac{1}{(b-\mathrm{z}_\star)(b-\mathrm{z}_\star-\lambda)}\right] = 0,$$

we conclude that $-1 < \mathrm{z}_\star{}'(\lambda) < 0$. In particular, $\mathrm{z}_\star$ is monotone and has limits as $\lambda \downarrow 0$ and $\lambda \uparrow \lambda_0$. We also have continuous differentiability of $\mathrm{z}_\star$ for $\lambda > \lambda_0$. Now, supposing $\lambda_0 \in (0, \underline{\alpha} + \underline{\beta})$, we show that $\mathrm{z}_\star$ is continuous at λ_0. Letting $\lambda \uparrow \lambda_0$ in (2.5.34), we obtain

$$\partial_z F(\lim_{\lambda\uparrow\lambda_0} \mathrm{z}_\star(\lambda), \lambda_0) = 0. \tag{2.5.36}$$

Since the minimizer occurs at the boundary when $\lambda = \lambda_0$, we deduce from (2.5.36) that $\lim_{\lambda\uparrow\lambda_1} \mathrm{z}_\star(\lambda) = -\underline{\alpha}$ and $\lim_{\lambda\uparrow\lambda_2} \mathrm{z}_\star(\lambda) = \underline{\beta} - \lambda_2$ when $\lambda_0 = \lambda_1$ and $\lambda_0 = \lambda_2$, respectively.

Since $\mathrm{z}_\star(\lambda) \in [-\underline{\alpha}, \underline{\beta} - \lambda]$, we have $\lim_{\lambda\uparrow\underline{\alpha}+\underline{\beta}} \mathrm{z}_\star(\lambda) = -\underline{\alpha}$. Set $\mathrm{z}_\star(0) = \lim_{\lambda\downarrow 0} \mathrm{z}_\star(\lambda)$. To calculate this limit, we consider several cases. If $\lambda_0 > 0$ then we can let $\lambda \downarrow 0$ in (2.5.34) and obtain

$$0 = \partial_z F(\mathrm{z}_\star(0), 0) = -s\,\mathrm{E}\left[\frac{1}{(a+\mathrm{z}_\star(0))^2}\right] + t\,\mathrm{E}\left[\frac{1}{(b-\mathrm{z}_\star(0))^2}\right] = \partial_z g_{\mathrm{z}_\star(0)}(s,t),$$

which implies $\mathrm{z}_\star(0) = \zeta$. If $\lambda_1 = 0$ then $\partial_z F(-\underline{\alpha}, 0) = \partial_z g_{-\underline{\alpha}}(s,t) \geqslant 0$ and if $\lambda_2 = 0$ then $\partial_z F(\underline{\beta}, 0) = \partial_z g_{\underline{\beta}}(s,t) \leqslant 0$. Hence, we get $\zeta = -\underline{\alpha} = \mathrm{z}_\star(0)$ and $\zeta = \underline{\beta} = \mathrm{z}_\star(0)$, respectively. $\square$

We omit the proof of Lemma 2.5.6 which is similar to that of Lemma 2.5.5.

Lemma 2.5.7. *For each $s, t > 0$, $\mathbf{L}_{s,t}$ is continuously differentiable on $[0, \underline{\alpha}+\underline{\beta})$ and $\mathbf{L}'_{s,t}(0) = g(s,t)$. Furthermore, $\mathbf{L}'_{s,t}$ is continuously differentiable on $(0, \underline{\alpha}+\underline{\beta}) \setminus \{\lambda_0\}$ and $\mathbf{L}''_{s,t} > 0$. The same statements also hold for $\mathbb{L}_{s,t}$.*

Proof. Let us write L for $\mathbf{L}_{s,t}$ and $F = F(z,\lambda)$ be given by (2.5.31). Using Lemma 2.5.5, we compute

$$L'(\lambda) = \partial_z F(\mathrm{z}_\star, \lambda)\,\mathrm{z}_\star{}'(\lambda) + \partial_\lambda F(\mathrm{z}_\star, \lambda) = s\,\mathrm{E}\left[\frac{1}{a+\mathrm{z}_\star+\lambda}\right] + t\,\mathrm{E}\left[\frac{1}{b-\mathrm{z}_\star-\lambda}\right] \tag{2.5.37}$$

for $0 < \lambda < \lambda_0$. Differentiating again, we obtain

$$L''(\lambda) = \partial_z\partial_\lambda F(\mathrm{z}_\star, \lambda)\,\mathrm{z}_\star{}'(\lambda) + \partial_\lambda^2 F(\mathrm{z}_\star, \lambda) = \frac{\partial_z^2 F(\mathrm{z}_\star,\lambda)\partial_\lambda^2 F(\mathrm{z}_\star,\lambda) - \partial_z\partial_\lambda F(\mathrm{z}_\star,\lambda)^2}{\partial_z^2 F(\mathrm{z}_\star,\lambda)} > 0,$$

where the inequality comes from $\partial_z^2 F(z_\star, \lambda) > \partial_\lambda \partial_z F(z_\star, \lambda)$ and $\partial_\lambda^2 F = \partial_\lambda \partial_z F$. For $\lambda > \lambda_1$,

$$L'(\lambda) = s\,\mathrm{E}\left[\frac{1}{a - \underline{\alpha} + \lambda}\right] + t\,\mathrm{E}\left[\frac{1}{b + \underline{\alpha} - \lambda}\right] \tag{2.5.38}$$

$$L''(\lambda) = -s\,\mathrm{E}\left[\frac{1}{(a - \underline{\alpha} + \lambda)^2}\right] + t\,\mathrm{E}\left[\frac{1}{(b + \underline{\alpha} - \lambda)^2}\right] > \partial_z F(-\underline{\alpha}, \lambda) > 0. \tag{2.5.39}$$

Also, for $\lambda > \lambda_2$,

$$L'(\lambda) = s\,\mathrm{E}\left[\frac{1}{a + \underline{\beta} - \lambda}\right] + t\,\mathrm{E}\left[\frac{1}{b - \underline{\beta} + \lambda}\right] \tag{2.5.40}$$

$$L''(\lambda) = s\,\mathrm{E}\left[\frac{1}{(a + \underline{\beta} - \lambda)^2}\right] - t\,\mathrm{E}\left[\frac{1}{(b - \underline{\beta} + \lambda)^2}\right] > -\partial_z F(\underline{\beta} - \lambda, \lambda) > 0. \tag{2.5.41}$$

We have verified that L is continuously differentiable on $(0, \underline{\alpha} + \underline{\beta}) \smallsetminus \{\lambda_0\}$ and L' is increasing.

We next note that L is also continuously differentiable at λ_0 when $\lambda_0 \in (0, \underline{\alpha} + \underline{\beta})$, for which it suffices to check that the left and right limits of L' at λ_0 match. First, we consider the case $\lambda_1 \in (0, \underline{\alpha} + \underline{\beta})$. Then, as $\lambda \uparrow \lambda_1$, (2.5.37) tends to $s\,\mathrm{E}[(a - \underline{\alpha} + \lambda_1)^{-1}] + t\,\mathrm{E}[(b - \underline{\alpha} - \lambda_1)^{-1}]$, which equals the $\lambda \downarrow \lambda_1$ limit of (2.5.38). Now, suppose that $\lambda_2 \in (0, \underline{\alpha} + \underline{\beta})$. Then, as $\lambda \uparrow \lambda_2$, (2.5.37) tends to $s\,\mathrm{E}[(a + \underline{\beta})^{-1}] + t\,\mathrm{E}[(b - \underline{\beta})^{-1}]$, which is the same as

$$s\,\mathrm{E}\left[\frac{1}{a + \underline{\beta} - \lambda_2}\right] + t\,\mathrm{E}\left[\frac{1}{b - \underline{\beta} + \lambda_2}\right] + \partial_z F(\underline{\beta} - \lambda_2, \lambda_2) = s\,\mathrm{E}\left[\frac{1}{a + \underline{\beta} - \lambda_2}\right] + t\,\mathrm{E}\left[\frac{1}{b - \underline{\beta} + \lambda_2}\right],$$

the $\lambda \downarrow \lambda_2$ limit of (2.5.40).

We next calculate $L'(0) = \lim_{\lambda \downarrow 0} L'(\lambda)$. If $\lambda_0 > 0$ then $\lambda \downarrow 0$ limit of (2.5.37) gives

$$L'(0) = s\,\mathrm{E}\left[\frac{1}{a + \zeta}\right] + t\,\mathrm{E}\left[\frac{1}{b - \zeta}\right] = g(s,t).$$

In the cases $\lambda_1 = 0$ and λ_2 then $\zeta = -\underline{\alpha}$ and $\zeta = \underline{\beta}$, respectively. Hence, letting $\lambda \downarrow 0$ in (2.5.38) and (2.5.40), respectively, we still obtain $L'(0) = g(s,t)$.

The asserted properties of $\mathbb{L}$ are proved similarly. $\square$

Since $\mathbf{L}'_{s,t}$ increasing, $\mathbf{L}'_{s,t}(\lambda)$ has a limit (possibly ∞) as $\lambda \uparrow \underline{\alpha} + \underline{\beta}$, which we denote by $\mathbf{L}'_{s,t}(\underline{\alpha} + \underline{\beta})$. Similarly, let us write $\mathbb{L}'_{s,t}(\underline{\alpha} + \underline{\beta})$ for $\lim_{\lambda \uparrow \underline{\alpha} + \underline{\beta}} \mathbb{L}'_{s,t}(\lambda)$. The precise values of these limits are needed in the next section.

Corollary 2.5.8. *Fix $s, t > 0$.*

$$\mathbf{L}'_{s,t}(\underline{\alpha}+\underline{\beta}) = \begin{cases} s\,\mathrm{E}\left[\dfrac{1}{a-\underline{\alpha}}\right] + t\,\mathrm{E}\left[\dfrac{1}{b+\underline{\alpha}}\right] & \textit{if } -s\,\mathrm{E}\left[\dfrac{1}{(a-\underline{\alpha})(a+\underline{\beta})}\right] + t\,\mathrm{E}\left[\dfrac{1}{(b+\underline{\alpha})(b-\underline{\beta})}\right] \leqslant 0 \\ s\,\mathrm{E}\left[\dfrac{1}{a+\underline{\beta}}\right] + t\,\mathrm{E}\left[\dfrac{1}{b-\underline{\beta}}\right] & \textit{otherwise.} \end{cases}$$

$$\mathbb{L}'_{s,t}(\underline{\alpha}+\underline{\beta}) = \begin{cases} s\dfrac{\mathrm{E}\left[\dfrac{a+\underline{\beta}}{(a-\underline{\alpha})^2}\right]}{\mathrm{E}\left[\dfrac{a+\underline{\beta}}{a-\underline{\alpha}}\right]} + t\dfrac{\mathrm{E}\left[\dfrac{1}{b-\underline{\beta}}\right]}{\mathrm{E}\left[\dfrac{b+\underline{\alpha}}{b-\underline{\beta}}\right]} & \textit{if } -s\dfrac{\mathrm{E}\left[\dfrac{1}{(a-\underline{\alpha})^2}\right]}{\mathrm{E}\left[\dfrac{a+\underline{\beta}}{a-\underline{\alpha}}\right]} + t\dfrac{\mathrm{E}\left[\dfrac{1}{(b-\underline{\beta})^2}\right]}{\mathrm{E}\left[\dfrac{b+\underline{\alpha}}{b-\underline{\beta}}\right]} \leqslant 0 \\ s\dfrac{\mathrm{E}\left[\dfrac{1}{a-\underline{\alpha}}\right]}{\mathrm{E}\left[\dfrac{a+\underline{\beta}}{a-\underline{\alpha}}\right]} + t\dfrac{\mathrm{E}\left[\dfrac{b+\underline{\alpha}}{(b-\underline{\beta})^2}\right]}{\mathrm{E}\left[\dfrac{b+\underline{\alpha}}{b-\underline{\beta}}\right]} & \textit{otherwise.} \end{cases}$$

The next lemma establishes continuous differentiability of $\mathbf{J}_{s,t}(r)$ and $\mathbb{J}_{s,t}(r)$ and shows that these functions are linear in r for $r > \mathbf{L}'_{s,t}(\underline{\alpha}+\underline{\beta})$ and $r > \mathbb{L}'_{s,t}(\underline{\alpha}+\underline{\beta})$, respectively.

Lemma 2.5.9. *Fix $s, t > 0$. For each $r \geqslant g(s,t)$, there exists a unique $\lambda_\star(r) \in [0, \underline{\alpha}+\underline{\beta}]$ such that $\mathbf{J}_{s,t}(t) = \lambda_\star r - \mathbf{L}_{s,t}(\lambda_\star)$. Moreover, $\mathbf{J}_{s,t}$ is continuously differentiable and $\mathbf{J}'_{s,t}(r) = \lambda_\star(r)$ for $r \geqslant g(s,t)$. If $r > g(s,t)$, then $\lambda_\star > 0$. If $r \geqslant \mathbf{L}'_{s,t}(\underline{\alpha}+\underline{\beta})$ then $\lambda_\star = \underline{\alpha}+\underline{\beta}$, while if $r \in [g(s,t), \mathbf{L}'_{s,t}(\underline{\alpha}+\underline{\beta}))$ then $\mathbf{L}'_{s,t}(\lambda_\star) = r$. The same statements hold if we replace $\mathbf{J}_{s,t}$ and $\mathbf{L}_{s,t}$ with $\mathbb{J}_{s,t}$ and $\mathbb{L}_{s,t}$, respectively.*

Proof. We have $J(r) = \sup_{0<\lambda<\underline{\alpha}+\underline{\beta}}\{\lambda r - L(\lambda)\}$, where (L, J) pair refers to either $(\mathbf{L}_{s,t}, \mathbf{J}_{s,t})$ or $(\mathbb{L}_{s,t}, \mathbb{J}_{s,t})$. The λ-derivative of the function inside the supremum is $r - L'(\lambda)$. By Lemma 2.5.7, L' is continuous and increasing from $g(s,t)$ to the limit $L'(\underline{\alpha}+\underline{\beta})$ on $(0, \underline{\alpha}+\underline{\beta})$. It follows that the unique maximizer $\lambda_\star$ is at $\underline{\alpha}+\underline{\beta}$ if $r \geqslant L'(\underline{\alpha}+\underline{\beta})$ and at $(L')^{-1}(r)$, otherwise. In addition, $\lambda_\star$ is increasing and continuous on $[g(s,t), +\infty)$. Since L' is differentiable and has nonzero derivative for $\lambda \in (0, \underline{\alpha}+\underline{\beta}) \smallsetminus \lambda_0$, whenever $r \neq L'(\lambda_0)$, we have $J'(r) = \lambda_\star(r) + {\lambda_\star}'(r) r - L'(\lambda_\star)\,{\lambda_\star}'(r) = \lambda_\star(r)$. Then continuity of $\lambda_\star$ implies that J is continuosly differentiable for all $r \geqslant g(s,t)$ including $L'(\lambda_0)$ when $\lambda_0 \in (0, \underline{\alpha}+\underline{\beta})$. $\square$

Proof of Theorem 2.2.27. This theorem is included in the preceding lemma. $\square$

2.5.3 Left tail estimates

We now estimate the left tail in both the quenched and annealed settings. The first result shows that in the quenched case, the rate n large deviation rate function are trivial for deviations to the left of the shape function $g(s,t)$. This proof is based on the proof of [80, Theorem 4.1], which was adapted from an argument in [55].

Proof of Lemma 2.2.22. First, fix $s, t, \epsilon > 0$ and rational. Take $m \in \mathbb{N}$ large enough that $m^{-1}\mathbb{E}\, G(\lfloor ms \rfloor, \lfloor mt \rfloor) \geqslant g(s,t) - \frac{\epsilon}{2}$. We coarse grain the lattice into pairwise disjoint translates of the set $\{1, \ldots, \lfloor ms \rfloor\} \times \{1, \ldots, \lfloor mt \rfloor\}$. Toward this end, define

$$A_{a,b}^{k,\ell} = \{1+a, \ldots, a+k\} \times \{1+b, \ldots, \ell+b\}, \qquad\qquad B_i^j = A_{(j+i)\lfloor ms \rfloor, j\lfloor mt \rfloor}^{\lfloor mx \rfloor, \lfloor my \rfloor}.$$

Take n large and let $L = \lfloor \frac{n}{m} - \lfloor \sqrt{n} \rfloor - 2 \rfloor$. For each such $k \leqslant \lfloor \sqrt{n} \rfloor$, define a diagonal by $D_k = \bigcup_{j=0}^{L} B_k^j$. We observe that the passage time from the bottom left corner of B_i^j to the top-right corner of B_i^j, $G_{i,j} \equiv G(\lfloor ms \rfloor, \lfloor mt \rfloor) \circ \tau_{(i+j)\lfloor ms \rfloor, j\lfloor mt \rfloor}$, has the same distribution as $G_{0,0}$ under $\mathbb{P}$. Moreover, if $(i_1, j_1) \neq (i_2, j_2)$, then $B_{i_1}^{j_1} \bigcap B_{i_2}^{j_2} = \varnothing$ and consequently $\{G_{i,j}\}_{i,j \geqslant 0}$ forms an independent family under $\mathbf{P}_{\mathbf{a},\mathbf{b}}$.

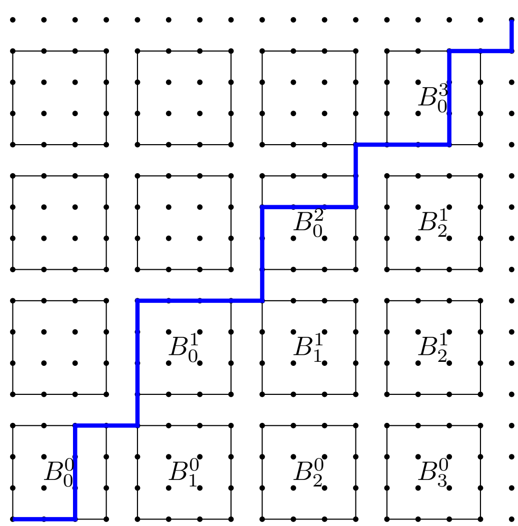


Figure 11: A path passing through the bottom-left and top-right vertices of B_0^j for each j.

Denote by Π_k the collection of paths from $(1,1)$ to $(\lfloor ns\rfloor, \lfloor nt\rfloor)$ passing through the bottom-left and top-right vertices of B_k^j for each j. See Figure 11. We have

$$G(\lfloor ns\rfloor, \lfloor nt\rfloor) \geqslant \max_{k\leqslant\lfloor\sqrt{n}\rfloor} \max_{\pi\in\Pi_k} \sum_{(i,j)\in\pi} W(i,j) \geqslant \max_{k\leqslant\lfloor\sqrt{n}\rfloor} \sum_{j\leqslant L} G_{k,j}.$$

It follows that

$$\begin{aligned}\mathbf{P}_{\mathbf{a},\mathbf{b}}\left(n^{-1}G(\lfloor ns\rfloor, \lfloor nt\rfloor) \leqslant (g(s,t)-\epsilon)\right) &\leqslant \mathbf{P}_{\mathbf{a},\mathbf{b}}\left(\max_{k\leqslant\lfloor\sqrt{n}\rfloor} n^{-1}\sum_{j=0}^{L} G_{k,j} \leqslant g(s,t)-\epsilon\right)\\ &= \prod_{k=0}^{\lfloor\sqrt{n}\rfloor} \mathbf{P}_{\mathbf{a},\mathbf{b}}\left(n^{-1}\sum_{j=0}^{L} G_{k,j} \leqslant g(s,t)-\epsilon\right).\end{aligned}$$

Now, fix $\lambda > 0$ sufficiently small that $C \equiv \lambda m\frac{\epsilon}{2} - \frac{\lambda^2}{2}\,\mathbb{E}\,G_{0,0}^2 > 0$ and $\lambda\,\mathbb{E}\,G_{0,0} - \frac{\lambda^2}{2}\,\mathbb{E}\,G_{0,0}^2 < 1$ and notice that $\mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{-\lambda G_{j,k}}\right] = \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{-\lambda G_{0,0}}\right] \circ \tau_{(j+k)\lfloor ms\rfloor, k\lfloor mt\rfloor}$. The ergodic theorem then implies that the following limit holds μ almost surely:

$$\lim_{L\to\infty} \frac{1}{L}\sum_{j=0}^{L} \log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{-\lambda G_{k,j}}\right] = \mathrm{E}\left[\log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{-\lambda G_{0,0}}\right]\right].$$

Jensen's inequality gives $\mathrm{E}\left[\log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{-\lambda G_{0,0}}\right]\right] \leqslant \log \mathbb{E}\left[e^{-\lambda G_{0,0}}\right] < -\lambda\,\mathbb{E}\,G_{0,0} + \frac{\lambda^2}{2}\,\mathbb{E}\,G_{0,0}^2$. By the exponential Markov inequality and independence under $\mathbf{P}_{\mathbf{a},\mathbf{b}}$, we have

$$\frac{1}{L}\log \mathbf{P}_{\mathbf{a},\mathbf{b}}\left(\sum_{j=0}^{L} G_{k,j} < n(g(s,t)-\epsilon)\right) \leqslant \frac{1}{L}\left(\sum_{j=0}^{L} \log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{-\lambda G_{j,k}}\right] + \lambda n(g(s,t)-\epsilon)\right).$$

Recalling that $L^{-1}n \to m$ as $n \to \infty$, and our assumption that $\mathbb{E}\,G_{0,0} > m(g(s,t) - \frac{\epsilon}{2})$, it follows that $\limsup_{L\to\infty} L^{-1}\log \mathbf{P}_{\mathbf{a},\mathbf{b}}\left(\sum_{j=0}^{L} G_{k,j} < n(g(s,t)-\epsilon)\right) \leqslant -\lambda m\frac{\epsilon}{2} + \frac{\lambda^2}{2}\,\mathbb{E}\,G_{0,0}^2 = -C$ almost surely. Therefore, for each k there exists a random N_k so that for $n \geqslant N_k$

$$\mathbf{P}_{\mathbf{a},\mathbf{b}}\left(\sum_{j=0}^{L} G_{k,j} < n(g(s,t)-\epsilon)\right) \leqslant \exp\left\{-n\frac{C}{2m}\right\}.$$

For any fixed K and $n \geqslant \max_{k\leqslant K} N_k$, we see that $\mathbb{P}$ almost surely we have

$$-\frac{1}{n}\log \mathbf{P}_{\mathbf{a},\mathbf{b}}\left(n^{-1}G(\lfloor ns\rfloor, \lfloor nt\rfloor) \leqslant (g(s,t)-\epsilon)\right) \geqslant \sum_{k=0}^{K} -\frac{1}{n}\log \mathbf{P}_{\mathbf{a},\mathbf{b}}\left(n^{-1}\sum_{j=0}^{L} G_{k,j} \leqslant g(s,t)-\epsilon\right)$$

$$\geqslant K\frac{C}{2m}.$$

Sending $n \to \infty$ and then $K \to \infty$ gives the result for fixed $s, t, \epsilon > 0$. For the general result, we work on the μ almost sure set where the result holds simultaneously for all rational $s, t, \epsilon > 0$. Take $s, t, \epsilon > 0$ and $s_1 < s$ and $t_1 < t$ rational with the property that $\epsilon - g(s,t) + g(s_1,t_1) > \epsilon_1 > 0$ for rational ϵ_1. This is possible by continuity of g. The result follows from observing that

$$\mathbf{P}_{\mathbf{a},\mathbf{b}}\left(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor) \leqslant g(s,t) - \epsilon\right) \leqslant \mathbf{P}_{\mathbf{a},\mathbf{b}}\left(n^{-1}G(\lfloor ns_1 \rfloor, \lfloor nt_1 \rfloor) \leqslant g(s_1,t_1) - \epsilon_1\right). \qquad \square$$

Corollary 2.5.10. *μ a.s. for $s, t, \lambda > 0$, $\lim_{n\to\infty} n^{-1} \log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[\exp\{-\lambda G(\lfloor ns \rfloor, \lfloor nt \rfloor)\}\right] = -\lambda g(s,t)$.*

Essentially the same argument as in Lemma 2.2.22 restricted to a single diagonal D_0 (so that the last passage times on B_0^j are i.i.d. under $\mathbb{P}$) shows that for $r \in (0, g(s,t))$, we have

$$\liminf_{n\to\infty} -n^{-1} \log \mathbb{P}\left(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor) \leqslant r\right) > 0.$$

To show that n is the correct rate for certain left tail large deviations, we need to show that the corresponding limsup is finite for some $r \in (0, g(s,t))$. We begin by considering the natural mechanism for these deviations, which we stated previously in Section **??** as Lemma 2.2.24.

Proof of Lemma 2.2.24. We may assume without loss of generality that $\{\nu_1 \in \mathcal{M}^\alpha, \nu_2 \in \mathcal{M}^\beta : g_{\nu_1,\nu_2}(s,t) \in (x,y)\} \neq \varnothing$ since the right hand side is infinite otherwise. Fix a pair ν_1, ν_2 from this set and introduce the notation

$$A_n = \{n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor) \in (x,y)\}, \qquad \frac{d\nu_1}{d\alpha}(a) = \varphi(a), \qquad \frac{d\nu_2}{d\beta}(b) = \psi(b).$$

Since A_n is measurable with respect to $\sigma\left(W(i,j) : 1 \leqslant i \leqslant \lfloor ns \rfloor, 1 \leqslant j \leqslant \lfloor nt \rfloor\right)$, we see that

$$\mathbb{P}_{\alpha,\beta}(A_n) = \mathrm{E}_{\alpha,\beta}\left[\mathbf{P}_{\mathbf{a},\mathbf{b}}(A_n)\right] \geqslant \mathrm{E}_{\alpha,\beta}\left[\mathbf{P}_{\mathbf{a},\mathbf{b}}(A_n) \prod_{i=1}^{\lfloor ns \rfloor} 1_{\{\varphi(a_i)>0\}} \prod_{j=1}^{\lfloor nt \rfloor} 1_{\{\psi(b_j)>0\}}\right]$$

$$= \mathrm{E}_{\nu_1,\nu_2}\left[\mathbf{P}_{\mathbf{a},\mathbf{b}}(A_n)\prod_{i=1}^{\lfloor ns\rfloor}\varphi(a_i)^{-1}\prod_{j=1}^{\lfloor nt\rfloor}\psi(b_j)^{-1}\right].$$

Taking logs and applying Jensen's inequality shows that

$$\begin{aligned}
-\frac{1}{n}\log\mathbb{P}_{\alpha,\beta}(A_n) &\leqslant -\frac{1}{n}\log\mathrm{E}_{\nu_1,\nu_2}\left[\mathbf{P}_{\mathbf{a},\mathbf{b}}(A_n)\prod_{i=1}^{\lfloor ns\rfloor}\varphi(a_i)^{-1}\prod_{j=1}^{\lfloor nt\rfloor}\psi(b_j)^{-1}\right]\\
&\leqslant \frac{1}{n\,\mathbb{P}_{\nu_1,\nu_2}(A_n)}\mathrm{E}_{\nu_1,\nu_2}\left[\mathbf{P}_{\mathbf{a},\mathbf{b}}(A_n)\left(\sum_{i=1}^{\lfloor ns\rfloor}\log\varphi(a_i)+\sum_{j=1}^{\lfloor nt\rfloor}\log\psi(b_j)\right)\right]-\frac{1}{n}\log\mathbb{P}_{\nu_1,\nu_2}(A_n).
\end{aligned}$$

Note that for any measures ν_1,ν_2, we have $g_{\nu_1,\nu_2}(s,t)>0$, so we have not divided by zero above. The last term tends to zero because $\mathbb{P}_{\nu_1,\nu_2}(A_n)\to 1$ as $n\to\infty$. For the remaining term, we note that

$$\begin{aligned}
\mathrm{E}_{\nu_1,\nu_2}\left[\mathbf{P}_{\mathbf{a},\mathbf{b}}(A_n)\left(\sum_{i=1}^{\lfloor ns\rfloor}\log\varphi(a_i)+\sum_{j=1}^{\lfloor nt\rfloor}\log\psi(b_j)\right)\right] &= \mathrm{E}_{\nu_1,\nu_2}\left[\left(\sum_{i=1}^{\lfloor ns\rfloor}\log\varphi(a_i)+\sum_{j=1}^{\lfloor nt\rfloor}\log\psi(b_j)\right)\right]\\
&\quad -\mathrm{E}_{\nu_1,\nu_2}\left[\mathbf{P}_{\mathbf{a},\mathbf{b}}(A_n^c)\log\left(\prod_{i=1}^{\lfloor ns\rfloor}\varphi(a_i)\prod_{j=1}^{\lfloor nt\rfloor}\psi(b_j)\right)\right]\\
&= \lfloor ns\rfloor\,\mathrm{H}(\nu_1|\alpha)+\lfloor nt\rfloor\,\mathrm{H}(\nu_2|\beta)\\
&\quad -\mathrm{E}_{\alpha,\beta}\left[\mathbf{P}_{\mathbf{a},\mathbf{b}}(A_n^c)\prod_{i=1}^{\lfloor ns\rfloor}\varphi(a_i)\prod_{j=1}^{\lfloor nt\rfloor}\psi(b_j)\log\left(\prod_{i=1}^{\lfloor ns\rfloor}\varphi(a_i)\prod_{j=1}^{\lfloor nt\rfloor}\psi(b_j)\right)\right]
\end{aligned}$$

But $x\log x\geqslant -\frac{1}{e}$ and $\mathbf{P}_{\mathbf{a},\mathbf{b}}(A_n^c)\in[0,1]$ so the last term is bounded above by a constant. Dividing by n and taking $\limsup_{n\to\infty}$, then optimizing over ν_1,ν_2 gives the result. □

To show that the annealed model has non-trivial rate n large deviations to the left of the shape function, it suffices to show that there exists $\nu_1\in\mathcal{M}^\alpha$ with $g_{\nu_1,\beta}(s,t)<g_{\alpha,\beta}(s,t)$. The next lemma gives mild conditions under which this is the case.

Lemma 2.5.11. *Suppose that α is not degenerate and $\mathrm{E}^\alpha[a\log a]<\infty$. Then there exists ν_1 with $\mathrm{H}(\nu_1|\alpha)<\infty$ and $g_{\nu_1,\beta}(s,t)<g_{\alpha,\beta}(s,t)$.*

Proof. Define ν_1 by $\frac{d\nu_1}{d\alpha}(a) = aE[a]^{-1}$. Note that $\mathrm{H}(\nu_1|\alpha) < \infty$ by hypothesis. Let $\zeta \in [-\underline{\alpha}, \underline{\beta}]$ be such that $g_{\alpha,\beta}(s,t) = s\,\mathrm{E}\left[(a+\zeta)^{-1}\right] + t\,\mathrm{E}\left[(b-\zeta)^{-1}\right]$. Because $\alpha \neq \delta_c$ for any c, the Cauchy-Schwarz inequality gives $1 = \mathrm{E}\left[\sqrt{a+\zeta}\sqrt{a+\zeta}^{-1}\right]^2 < \mathrm{E}[a+\zeta]\,\mathrm{E}\left[(a+\zeta)^{-1}\right]$. Rearranging implies that $\mathrm{E}\left[a(a+\zeta)^{-1}\right] < \mathrm{E}[a]\,\mathrm{E}\left[(a+\zeta)^{-1}\right]$. It then follows that

$$g_{\nu_1,\beta}(s,t) \leqslant s\,\mathrm{E}[a]^{-1}\,\mathrm{E}\left[\frac{a}{a+\zeta}\right] + t\,\mathrm{E}\left[\frac{1}{b-\zeta}\right] < s\,\mathrm{E}\left[\frac{1}{a+\zeta}\right] + t\,\mathrm{E}\left[\frac{1}{b-\zeta}\right] = g_{\alpha,\beta}(s,t). \qquad \square$$

We expect that the moment condition in the previous lemma is unnecessary.

2.5.4 Large deviation principle

We prove Theorem 2.2.18 by working with Legendre-Fenchel transforms and appealing to convex duality.

Lemma 2.5.12. *For all $s,t > 0$,*

$$\mathbf{J}^\star_{s,t}(\lambda) = \begin{cases} \mathbf{L}_{s,t}(\lambda) & \lambda \geqslant 0 \\ \infty & \lambda < 0 \end{cases}, \qquad \mathbb{J}^\star_{s,t}(\lambda) = \begin{cases} \mathbb{L}_{s,t}(\lambda) & \lambda \geqslant 0 \\ \infty & \lambda < 0 \end{cases}.$$

Proof. We give the proof of the result under $\mathbf{P}_{\mathbf{a},\mathbf{b}}$. The proof under $\mathbb{P}$ is similar. Recall the regularity properties of $\mathbf{J}_{s,t}(\cdot)$ proven in Proposition 2.5.14 in the appendix. The result for $\lambda < 0$ follows from the observation that $\mathbf{J}_{s,t}(r) = 0$ for $r \leqslant g(s,t)$. For all $\lambda > 0$, by the exponential Markov inequality we have

$$\frac{1}{n}\log \mathbf{P}_{\mathbf{a},\mathbf{b}}\left(G(\lfloor ns\rfloor, \lfloor nt\rfloor) \geqslant nr\right) \leqslant \frac{1}{n}\log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda G(\lfloor ns\rfloor,\lfloor nt\rfloor)}\right] - \lambda r.$$

Sending $n \to \infty$ gives $\lambda r - \mathbf{J}_{s,t}(r) \leqslant \mathbf{L}_{s,t}(\lambda)$ and taking $\sup_{r\in\mathbb{R}}$ implies $\mathbf{J}^\star_{s,t}(\lambda) \leqslant \mathbf{L}_{s,t}(\lambda)$. For the reverse inequality, we next consider the case $\lambda \in (0, \underline{\alpha} + \underline{\beta})$. Fix $M > 0$ and let $\{x_i\}_{i=0}^K$ be a partition of $[0, M]$. We observe that

$$\mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda G(\lfloor ns\rfloor,\lfloor nt\rfloor)}\right] = \sum_{i=1}^K \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda G(\lfloor ns\rfloor,\lfloor nt\rfloor)} 1_{(x_{i-1},x_i]}(n^{-1}G(\lfloor ns\rfloor,\lfloor nt\rfloor))\right]$$

$$+ \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda G(\lfloor ns \rfloor, \lfloor nt \rfloor)} 1_{(M,\infty)}(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor))\right].$$

Consequently, we see that

$$\begin{aligned}\frac{1}{n}\log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda G(\lfloor ns \rfloor, \lfloor nt \rfloor)}\right] \leqslant \max\Bigg\{ &\max_{0\leqslant i\leqslant K}\{\lambda x_i + \frac{1}{n}\log \mathbf{P}_{\mathbf{a},\mathbf{b}}\left(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor) \geqslant x_{i-1}\right)\}, \\ &\frac{1}{n}\mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda G(\lfloor ns \rfloor, \lfloor nt \rfloor)} 1_{(M,\infty)}(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor))\right]\Bigg\} + \frac{K+1}{n}\end{aligned}$$

Take $\limsup_{n\to\infty}$ then $K \to \infty$. Using continuity of $r \mapsto \mathbf{J}_{s,t}(r)$, we see that

$$\mathbf{L}_{s,t}(\lambda) \leqslant \max_{0\leqslant r\leqslant M}\{\lambda r - \mathbf{J}_{s,t}(r)\} \vee \limsup_{n\to\infty}\frac{1}{n}\log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda G(\lfloor ns \rfloor, \lfloor nt \rfloor)} 1_{(M,\infty)}(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor))\right].$$

Let $p, q > 1$ be such that $p^{-1} + q^{-1} = 1$ and $p\lambda < \underline{\alpha} + \underline{\beta}$. Then

$$\begin{aligned}\frac{1}{n}\log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda G(\lfloor ns \rfloor, \lfloor nt \rfloor)} 1_{(M,\infty)}(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor))\right] &\leqslant \frac{1}{pn}\log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda p G(\lfloor ns \rfloor, \lfloor nt \rfloor)}\right] \\ &+ \frac{1}{qn}\log \mathbf{P}_{\mathbf{a},\mathbf{b}}\left(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor \geqslant M\right).\end{aligned}$$

From this, we see that there exist deterministic constants C_1, C_2 such that

$$\limsup_{n\to\infty}\frac{1}{n}\log \mathbf{E}_{\mathbf{a},\mathbf{b}}\left[e^{\lambda G(\lfloor ns \rfloor, \lfloor nt \rfloor)} 1_{(M,\infty)}(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor))\right] \leqslant C_1 - C_2\,\mathbf{J}_{s,t}(M).$$

Recall that $\lambda r \leqslant \mathbf{L}_{s,t}(\lambda) + \mathbf{J}_{s,t}(r)$, so that as $M \to \infty$, $\mathbf{J}_{s,t}(M) \to \infty$. Since $\max_{r\leqslant M}\{\lambda r - \mathbf{J}_{s,t}(r)\} \leqslant \mathbf{J}^\star_{s,t}(\lambda)$, it follows that we have $\mathbf{L}_{s,t}(\lambda) \leqslant \mathbf{J}^\star_{s,t}(\lambda)$.

Next, we turn to the case $\lambda = \underline{\alpha} + \underline{\beta}$. We observe that as $\lambda \uparrow \underline{\alpha} + \underline{\beta}$, $\mathbf{L}_{s,t}(\lambda) \uparrow \mathbf{L}_{s,t}(\underline{\alpha} + \underline{\beta})$. Suppose that $\mathbf{L}_{s,t}(r) < \infty$. Fix $\epsilon > 0$ and take $\lambda < \underline{\alpha} + \underline{\beta}$ such that $\sup_{r\in\mathbb{R}}\{\lambda r - \mathbf{J}_{s,t}(r)\} = \mathbf{L}_{s,t}(\lambda) \geqslant \mathbf{L}_{s,t}(\underline{\alpha} + \underline{\beta}) - 2\epsilon$. Then there exists $r > 0$ so that $\lambda r - \mathbf{J}_{s,t}(r) \geqslant \mathbf{L}_{s,t}(\underline{\alpha} + \underline{\beta}) - \epsilon$. Since $(\underline{\alpha} + \underline{\beta})r > \lambda r$, it follows that $\mathbf{J}^\star_{s,t}(\underline{\alpha} + \underline{\beta}) \geqslant \mathbf{L}_{s,t}(\underline{\alpha} + \underline{\beta}) - \epsilon$. The case $\mathbf{L}_{s,t}(\underline{\alpha} + \underline{\beta}) = \infty$ is similar.

Finally, we consider the case $\lambda > \underline{\alpha} + \underline{\beta}$, where $\mathbf{L}_{s,t}(\lambda) = \infty$. For each (i, j), we eventually have $G(\lfloor ns \rfloor, \lfloor nt \rfloor) \geqslant W(i, j)$. This implies that for all (i, j), $\mathbf{J}_{s,t}(r) \leqslant (a_i + b_j)r1_{\{r\geqslant 0\}}$ and therefore μ almost surely, $\mathbf{J}_{s,t}(r) \leqslant (\underline{\alpha} + \underline{\beta})r1_{\{r\geqslant 0\}}$. Taking Legendre-Fenchel transforms of this inequality shows that $\mathbf{J}^\star_{s,t}(\lambda) = \infty$. $\square$

Proof of Theorem 2.2.18. Proposition 2.5.14 shows that $r \mapsto \mathbf{J}_{s,t}(r)$ and $r \mapsto \mathbb{J}_{s,t}(r)$ are real valued convex functions on $\mathbb{R}$. The result follows from taking Legendre-Fenchel transforms of the expressions in the previous lemma [79, Theorem 12.2]. $\square$

Proof of Theorem 2.2.23. Fix an open set $O \subset \mathbb{R}$

1. If $O \subset (-\infty, g(s,t))$ then there is nothing to prove by Lemma 2.2.22.

2. If $g(s,t) \in O$, then

$$\limsup_{n\to\infty} -n^{-1} \log \mathbf{P}_{\mathbf{a},\mathbf{b}} \left(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor) \in O\right) = 0 = \inf_{r \in O} I_{s,t}(r)$$

3. If $O \cap (g(s,t), \infty) \neq \varnothing$, then $O \cap (g(s,t), \infty)$ contains an interval (r_0, r_1). Note that

$$\begin{aligned}\mathbf{P}_{\mathbf{a},\mathbf{b}} \left(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor) \in O\right) &\geqslant \mathbf{P}_{\mathbf{a},\mathbf{b}} \left(G(\lfloor ns \rfloor, \lfloor nt \rfloor) \in (r_0, r_1)\right) \\ &= \mathbf{P}_{\mathbf{a},\mathbf{b}} \left(G(\lfloor ns \rfloor, \lfloor nt \rfloor) \geqslant r_0\right) - \mathbf{P}_{\mathbf{a},\mathbf{b}} \left(G(\lfloor ns \rfloor, \lfloor nt \rfloor) \geqslant r_1\right)\end{aligned}$$

   Lemma 2.5.9 shows that $\mathbf{J}_{s,t}(r)$ is strictly increasing for $r > g(s,t)$, which implies that

$$\limsup_{n\to\infty} -n^{-1} \log \mathbf{P}_{\mathbf{a},\mathbf{b}} \left(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor) \in O\right) \leqslant \mathbf{J}_{s,t}(r_0).$$

   Let $r_n \in O \cap (g(s,t), \infty)$ be a sequence with $r_n \downarrow r_\infty = \inf\{x : x \in O \cap (g(s,t), \infty)\}$. Then because $\mathbf{J}_{s,t}(r)$ is continuous and non-decreasing, we see that

$$\limsup_{n\to\infty} -n^{-1} \log \mathbf{P}_{\mathbf{a},\mathbf{b}} \left(n^{-1}G(\lfloor ns \rfloor, \lfloor nt \rfloor) \in O\right) \leqslant \mathbf{J}_{s,t}(r_\infty) = \inf_{r \in O \cap (g(s,t),\infty)} \mathbf{I}_{s,t}(r) = \inf_{r \in O} \mathbf{I}_{s,t}(r).$$

The upper bound follows from the regularity of $\mathbf{J}_{s,t}$, Theorem 2.2.18 and Lemma 2.2.22. $\square$

2.5.5 Relative entropy and the rate functions

We now turn to the proof of Theorem 2.2.26. Our argument proving this result is purely convex analytic and does not show the probabilistic interpretation mentioned before the statement of the theorem. We begin with a technical lemma.

Lemma 2.5.13. *For $r > 0$, the map $(\alpha, \beta) \mapsto \mathbf{I}_{s,t}^{\alpha,\beta}(r)$ is convex on $\mathcal{M}_1(\mathbb{R}_+)^2$.*

Proof. Using (2.2.14), one can check that $(\alpha, \beta) \mapsto g_{\alpha,\beta}(s,t)$ is concave on $\mathcal{M}(\mathbb{R}_+)^2$. Thus, $\{(\alpha, \beta) : g_{\alpha,\beta}(s,t) \geqslant r\}$ is convex. Similarly,

$$F(\alpha, \beta) = \sup_{\substack{\lambda \in (0, \underline{\alpha}+\underline{\beta}) \\ z \in (-\underline{\alpha}, \underline{\beta}-\lambda)}} \left\{ \lambda r - s\,\mathrm{E}^{\alpha}\left[\log\left(\frac{a+z+\lambda}{a+z}\right)\right] - t\,\mathrm{E}^{\beta}\left[\log\left(\frac{b-z}{b-z-\lambda}\right)\right]\right\}$$

is convex on $\mathcal{M}_1(\mathbb{R}_+)^2$. Then we see from (2.2.22) that $(\alpha, \beta) \mapsto \mathbf{I}_{s,t}^{\alpha,\beta}(r)$ is convex on $\mathcal{M}(\mathbb{R}^+)^2$. $\square$

Proof of Theorem 2.2.26. Theorem 2.2.18 and the variational characterization of relative entropy, [77, Theorem 5.4], imply that for $r > g(s,t)$,

$$\begin{aligned}
\mathbb{J}_{s,t}^{\alpha,\beta}(r) &= \sup_{\substack{\lambda \in (0, \underline{\alpha}+\underline{\beta}) \\ z \in (-\underline{\alpha}, \underline{\beta}-\lambda)}} \left\{ \lambda r - s \log \mathrm{E}^{\alpha}\left[\frac{a+z+\lambda}{a+z}\right] - t \log \mathrm{E}^{\beta}\left[\frac{b-z}{b-z-\lambda}\right]\right\} \\
&= \sup_{\substack{\lambda \in (0, \underline{\alpha}+\underline{\beta}) \\ z \in (-\underline{\alpha}, \underline{\beta}-\lambda)}} \inf_{\substack{\nu_1 \in \mathcal{M}^{\alpha} \\ \nu_2 \in \mathcal{M}^{\beta}}} \left\{ \lambda r - s\,\mathrm{E}^{\nu_1}\left[\log\left(\frac{a+z+\lambda}{a+z}\right)\right] - t\,\mathrm{E}^{\nu_2}\left[\log\left(\frac{b-z}{b-z-\lambda}\right)\right] \right. \\
&\qquad\qquad \left. + s\,\mathrm{H}(\nu_1|\alpha) + t\,\mathrm{H}(\nu_2|\beta) \right\} \\
&\leqslant \inf_{\substack{\nu_1 \in \mathcal{M}^{\alpha} \\ \nu_2 \in \mathcal{M}^{\beta}}} \sup_{\substack{\lambda \in (0, \underline{\alpha}+\underline{\beta}) \\ z \in (-\underline{\alpha}, \underline{\beta}-\lambda)}} \left\{ \lambda r - s\,\mathrm{E}^{\nu_1}\left[\log\left(\frac{a+z+\lambda}{a+z}\right)\right] - t\,\mathrm{E}^{\nu_2}\left[\log\left(\frac{b-z}{b-z-\lambda}\right)\right] \right. \\
&\qquad\qquad \left. + s\,\mathrm{H}(\nu_1|\alpha) + t\,\mathrm{H}(\nu_2|\beta) \right\}.
\end{aligned}$$

Note that if $\nu_1 \ll \alpha$, it must be the case that $\underline{\nu}_1 \geqslant \underline{\alpha}$ and similarly, $\underline{\nu}_2 \geqslant \underline{\beta}$. It follows that we may extend the region in the inner supremum to obtain

$$\mathbb{J}_{s,t}^{\alpha,\beta}(r) \leqslant \inf_{\nu_1, \nu_2} \left\{ \mathbf{I}_{s,t}^{\nu_1,\nu_2}(r) + s\,\mathrm{H}(\nu_1|\alpha) + t\,\mathrm{H}(\nu_2|\beta) \right\}.$$

The map $(\nu_1, \nu_2) \mapsto \mathbf{I}_{s,t}^{\nu_1,\nu_2}(r) + s\,\mathrm{H}(\nu_1|\alpha) + t\,\mathrm{H}(\nu_2|\beta)$ is strictly convex on the convex set $\mathcal{M}^{\alpha} \times \mathcal{M}^{\beta}$ so at most one minimizing pair (ν_1, ν_2) exists. It therefore suffices to show that we have equality with the measures ν_1, ν_2 defined in the statement of the theorem. We argue this by cases.

A maximizing pair $\lambda_\star, \mathrm{z}_\star$ satisfying $\lambda_\star \in [0, \underline{\alpha} + \underline{\beta}], \mathrm{z}_\star \in [-\underline{\alpha}, \underline{\beta} - \lambda_\star]$ exist for the annealed right-tail rate function by Lemmas 2.5.6 and 2.5.9. ($\mathrm{z}_\star$ denotes $\mathrm{z}_\star(\lambda_\star)$ in the notation of Section 2.5.2. Also, by Corollary 2.5.4, $\mathrm{z}_\star(\underline{\alpha} + \underline{\beta}) = -\underline{\alpha}$). Note that $\lambda_\star = 0$ is impossible because $\mathbb{J}^{\alpha,\beta}_{s,t}(r) > 0$ by Lemma 2.5.9. If $\lambda_\star \in (0, \underline{\alpha} + \underline{\beta})$ and $\mathrm{z}_\star \in (-\underline{\alpha}, \underline{\beta} - \lambda_\star)$, then $\nu_1 \in \mathcal{M}^\alpha$ and $\nu_2 \in \mathcal{M}^\beta$ because their densities with respect to α and β are bounded. Taking derivatives in (2.2.23), we see that $\mathrm{z}_\star$ and $\lambda_\star$ solve

$$0 = s\,\mathrm{E}^{\nu_1}\left[\frac{1}{a + \mathrm{z}_\star} - \frac{1}{a + \mathrm{z}_\star + \lambda_\star}\right] + t\,\mathrm{E}^{\nu_2}\left[\frac{1}{b - \mathrm{z}_\star} - \frac{1}{b - \mathrm{z}_\star - \lambda_\star}\right] \tag{2.5.42}$$

$$0 = r - s\,\mathrm{E}^{\nu_1}\left[\frac{1}{a + \mathrm{z}_\star + \lambda_\star}\right] - t\,\mathrm{E}^{\nu_2}\left[\frac{1}{b - \mathrm{z}_\star - \lambda_\star}\right]. \tag{2.5.43}$$

These are precisely the first order conditions implying that

$$\mathbf{I}^{\nu_1,\nu_2}_{s,t}(r) = \lambda_\star r - s\,\mathrm{E}^{\nu_1}\left[\log\frac{a + \mathrm{z}_\star + \lambda_\star}{a + \mathrm{z}_\star}\right] - t\,\mathrm{E}^{\nu_2}\left[\log\frac{b - \mathrm{z}_\star}{b - \mathrm{z}_\star - \lambda_\star}\right].$$

The definition of relative entropy and a little algebra then show that

$$\mathbb{J}^{\alpha,\beta}_{s,t}(r) = \mathbf{I}^{\nu_1,\nu_2}_{s,t}(r) + s\,\mathrm{H}(\nu_1|\alpha) + t\,\mathrm{H}(\nu_2|\alpha).$$

The remaining cases are similar in that once we know that the extremizers are the same for $\mathbb{J}^{\alpha,\beta}_{s,t}(r)$ and $\mathbf{I}^{\nu_1,\nu_2}_{s,t}(r)$, the result follows. The necessary and sufficient conditions in Lemmas 2.5.5 and 2.5.6 show that ν_1 and ν_2 are well defined and that this equality continues to hold if $\lambda_\star < \underline{\alpha} + \underline{\beta}$ and $\mathrm{z}_\star = -\underline{\alpha}$ or $\mathrm{z}_\star = \underline{\beta} - \lambda_\star$. The only remaining case is $\lambda_\star = \underline{\alpha} + \underline{\beta}$ and $\mathrm{z}_\star = -\underline{\alpha}$. $\lambda_\star = \underline{\alpha} + \underline{\beta}$ is equivalent to $r \geqslant (\mathbb{L}^{\alpha,\beta}_{s,t})'(\underline{\alpha} + \underline{\beta})$. By Corollary 2.5.8, this condition implies that ν_1 and ν_2 are well defined and $(\mathbf{L}^{\nu_1,\nu_2}_{s,t})'(\underline{\alpha} + \underline{\beta}) = (\mathbb{L}^{\alpha,\beta}_{s,t})'(\underline{\alpha} + \underline{\beta})$. The result follows. $\square$

2.5.6 Scaling estimates

In this section, we prove the scaling estimates for the quenched and the annealed rate functions. See the discussion Section 2.5.2 for the notation below. If $c_1 < s/t < c_2$ we have $\partial_z g_\zeta(s,t) = 0$

and, therefore,

$$g_z(s,t) = g(s,t) + \partial_z^2 g_\zeta(s,t)(z-\zeta)^2/2 + o((z-\zeta)^2). \tag{2.5.44}$$

In fact, (2.5.44) holds for $s/t = c_1$ and $s/t = c_2$ as well provided that

$$\mathrm{E}\left[\frac{1}{(a-\underline{\alpha})^3}\right] < \infty, \qquad \mathrm{E}\left[\frac{1}{(b-\underline{\beta})^3}\right] < \infty; \tag{2.5.45}$$

that is, assuming that $\partial_z^2 g_z(s,t)$ has limits at the endpoints $-\underline{\alpha}$ and $\underline{\beta}$.

Proof of Theorem 2.2.28. For $\epsilon > 0$ sufficiently small, we have

$$\mathbf{I}'_{s,t}(r) = \lambda_\star(r), \qquad \mathbf{L}'_{s,t}(\lambda_\star(r)) = r \tag{2.5.46}$$

whenever $g(s,t) \leqslant r \leqslant g(s,t)+\epsilon$ by Lemma 2.5.9. We begin with the case $c_1 < s/t < c_2$. Then $\zeta \in (-\underline{\alpha}, \underline{\beta})$. We recall λ_1 and λ_2 defined in Lemma 2.5.5. Because $\partial_z F(-\underline{\alpha}, 0) = \partial_z g_{-\underline{\alpha}}(s,t) < 0$ and $\partial_z F(-\underline{\alpha}, 0) = \partial_z g_{-\underline{\alpha}}(s,t) > 0$, we conclude that $\lambda_1 > 0$ and $\lambda_2 > 0$. Hence,

$$\mathrm{z}_\star{}'(\lambda) = -\frac{\partial_\lambda \partial_z F(\mathrm{z}_\star, \lambda)}{\partial_z^2 F(\mathrm{z}_\star,\lambda)} = -\frac{s\,\mathrm{E}\left[\dfrac{1}{(a+\mathrm{z}_\star)(a+\mathrm{z}_\star+\lambda)^2}\right] + t\,\mathrm{E}\left[\dfrac{1}{(b-\mathrm{z}_\star)(b-\mathrm{z}_\star-\lambda)^2}\right]}{s\,\mathrm{E}\left[\dfrac{2a+2\,\mathrm{z}_\star+\lambda}{(a+\mathrm{z}_\star+\lambda)^2(a+\mathrm{z}_\star)^2}\right] + t\,\mathrm{E}\left[\dfrac{2b-2\,\mathrm{z}_\star-\lambda}{(b-\mathrm{z}_\star-\lambda)^2(b-\mathrm{z}_\star)^2}\right]}.$$

for $0 < \lambda < \lambda_1 \wedge \lambda_2$. Letting $\lambda \downarrow 0$ yields $\mathrm{z}_\star{}'(0^+) = -1/2$. It follows that $\mathrm{z}_\star(\lambda) = \zeta - \lambda/2 + o(\lambda)$ as $\lambda \downarrow 0$. We obtain $\mathbf{L}'_{s,t}(\lambda) = g_{\mathrm{z}_\star+\lambda}(s,t) = g(s,t) + \partial_z^2 g_\zeta(s,t)\lambda^2/8 + o(\lambda^2)$ as $\lambda \downarrow 0$. Then,

$$\mathbf{I}'_{s,t}(g(s,t)+\epsilon) = \frac{2\sqrt{2}}{\sqrt{\partial_z^2 g_\zeta(s,t)}}\epsilon^{1/2} + o(\epsilon^{1/2}),$$

and integrating gives

$$\mathbf{I}_{s,t}(g(s,t)+\epsilon) = \frac{4\sqrt{2}\epsilon^{3/2}}{3\sqrt{\partial_z^2 g_\zeta(s,t)}} + o(\epsilon^{3/2}) = \frac{4}{3}\frac{\epsilon^{3/2}}{\sqrt{s\,\mathrm{E}\left[\dfrac{1}{(a+\zeta)^3}\right] + t\,\mathrm{E}\left[\dfrac{1}{(b-\zeta)^3}\right]}} + o(\epsilon^{3/2}) \tag{2.5.47}$$

as $\epsilon \downarrow 0$. Now, suppose that $s/t \leqslant c_1$. Then $E[(a-\underline{\alpha})^{-2}] < \infty$, $\zeta = -\underline{\alpha}$ and $\mathrm{z}_\star = -\underline{\alpha}$. Under condition (2.5.45), when $c_1 = s/t$, $\mathbf{L}'_{s,t}(\lambda) = g_{-\underline{\alpha}+\lambda}(s,t) = g(s,t) + \partial_z^2 g_{-\underline{\alpha}}(s,t)\lambda^2/2 + o(\lambda^2)$

and we reach (2.5.47) multiplied with 1/2. If $c_1 > s/t$ then $\partial_z g_{-\underline{\alpha}}(s,t) > 0$ and we have $\mathbf{L}'_{s,t}(\lambda) = g_{-\underline{\alpha}+\lambda}(s,t) = g(s,t) + \partial_z g_{-\underline{\alpha}}(s,t)\lambda + o(\lambda)$. This leads to

$$\mathbf{I}_{s,t}(g(s,t)+\epsilon) = \frac{\epsilon^2}{2\partial_z g_{-\underline{\alpha}}(s,t)} + o(\epsilon^2) = \frac{1}{2}\frac{\epsilon^2}{-s\,\mathrm{E}\left[\dfrac{1}{(a-\underline{\alpha})^2}\right] + t\,\mathrm{E}\left[\dfrac{1}{(b+\underline{\alpha})^2}\right]} + o(\epsilon^2).$$

Analysis of the case $s/t \geqslant c_2$ is similar. $\square$

Proof of Theorem 2.2.29. In the case $c_1 < s/t < c_2$, Hölder's inequality gives

$$\begin{aligned}
\lim_{\lambda\downarrow 0} \mathrm{z}_\star{}'(\lambda) &= -\lim_{\lambda\downarrow 0} \frac{\partial_\lambda \partial_z F(\mathrm{z}_\star,\lambda)}{\partial_z^2 F(\mathrm{z}_\star,\lambda)} \\
&= -\frac{s\,\mathrm{E}\left[\dfrac{1}{(a+\zeta)^2}\right]\mathrm{E}\left[\dfrac{1}{a+\zeta}\right] + 2t\,\mathrm{E}\left[\dfrac{1}{(b-\zeta)^3}\right] - t\,\mathrm{E}\left[\dfrac{1}{(b-\zeta)^2}\right]\mathrm{E}\left[\dfrac{1}{b-\zeta}\right]}{2s\,\mathrm{E}\left[\dfrac{1}{(a+\zeta)^3}\right] + 2t\,\mathrm{E}\left[\dfrac{1}{(b-\zeta)^3}\right]} \\
&\leqslant -\frac{s\,\mathrm{E}\left[\dfrac{1}{(a+\zeta)^2}\right]\mathrm{E}\left[\dfrac{1}{a+\zeta}\right] + t\,\mathrm{E}\left[\dfrac{1}{(b-\zeta)^3}\right]}{2s\,\mathrm{E}\left[\dfrac{1}{(a+\zeta)^3}\right] + 2t\,\mathrm{E}\left[\dfrac{1}{(b-\zeta)^3}\right]}.
\end{aligned}$$

Hence, $\mathrm{z}_\star(\lambda) = \zeta + c\lambda + o(\lambda)$, where $c < 0$. We have

$$\begin{aligned}
\mathbb{L}'_{s,t}(\lambda) &= s\frac{\mathrm{E}\left[\dfrac{1}{a+\mathrm{z}_\star}\right]}{\mathrm{E}\left[\dfrac{a+\mathrm{z}_\star+\lambda}{a+\mathrm{z}_\star}\right]} + t\frac{\mathrm{E}\left[\dfrac{b-\mathrm{z}_\star}{(b-\mathrm{z}_\star-\lambda)^2}\right]}{\mathrm{E}\left[\dfrac{b-\mathrm{z}_\star}{b-\mathrm{z}_\star-\lambda}\right]} &\quad (2.5.48)\\
&= g_{\mathrm{z}_\star+\lambda}(s,t) + \lambda\left(s\,\mathrm{Var}\left[\frac{1}{a+\zeta}\right] + t\,\mathrm{Var}\left[\frac{1}{b-\zeta}\right]\right) + o(\lambda) &\quad (2.5.49)\\
&= g(s,t) + \lambda\left(s\,\mathrm{Var}\left[\frac{1}{a+\zeta}\right] + t\,\mathrm{Var}\left[\frac{1}{b-\zeta}\right]\right) + o(\lambda). &\quad (2.5.50)
\end{aligned}$$

Then, arguing as in the preceding proof, we obtain

$$\mathbb{J}_{s,t}(g(s,t)+\epsilon) = \frac{1}{2}\frac{\epsilon^2}{s\,\mathrm{Var}\left[\dfrac{1}{a+\zeta}\right] + t\,\mathrm{Var}\left[\dfrac{1}{b-\zeta}\right]} + o(\epsilon). \quad (2.5.51)$$

Now consider $s/t \leqslant c_1$. Then $\zeta = \mathrm{z}_\star = -\underline{\alpha}$ and (2.5.48) still holds. If $s/t = c_1$ subsequent arguments go through assuming (2.5.45). This condition is needed in step (2.5.49), which relies on (2.5.44) with $\zeta = -\underset{\sim}{\alpha}$. Hence, we have (2.5.51). If $s/t < c_1$ then the coefficient of λ in (2.5.50) has an additional term $\partial_z g_{-\underset{\sim}{\alpha}}(s,t) > 0$, which leads to

$$\mathbb{J}_{s,t}(g(s,t)+\epsilon) = \frac{1}{2}\frac{\epsilon^2}{-s\,\mathrm{E}\left[\dfrac{1}{a-\underset{\sim}{\alpha}}\right]^2 + t\,\mathrm{Var}\left[\dfrac{1}{b+\underset{\sim}{\alpha}}\right] + t\,\mathrm{E}\left[\dfrac{1}{(b+\underset{\sim}{\alpha})^2}\right]} + o(\epsilon).$$

The case $s/t \geqslant c_2$ is analyzed similarly. $\square$

2.5.7 Right tail rate functions and Lyapunov exponents

Proposition 2.5.14. *(a) μ-a.s., for $s,t > 0$ and $r \in \mathbb{R}$, there exists (nonrandom) $\mathbf{J}_{s,t}(r) \in [0,\infty)$ such that*

$$\lim_{n\to\infty} -\frac{1}{n}\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor ns\rfloor, \lfloor nt\rfloor) \geqslant nr) = \mathbf{J}_{s,t}(r). \tag{2.5.52}$$

(b) For all $s,t > 0$ and $r \in \mathbb{R}$, there exists $\mathbb{J}_{s,t}(r) \in [0,\infty)$ such that

$$\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(G(\lfloor ns\rfloor, \lfloor nt\rfloor) \geqslant nr) = \mathbb{J}_{s,t}(r). \tag{2.5.53}$$

(c) $\mathbf{J}$ and $\mathbb{J}$ are convex and homogeneous in (s,t,r), nonincreasing in (s,t) and nondecreasing in r.

Proof. Fix $r \in \mathbb{R}$ and $s,t \in \mathbb{N}$. For integers $0 \leqslant m < n$, define

$$X_{m,n} = -\log \mathbf{P}_{\tau_{ms}(\mathbf{a}),\tau_{mt}(\mathbf{b})}(G((n-m)s,(n-m)t) \geqslant (n-m)r).$$

We verify that $\{X_{m,n}\}$ satisfy the hypotheses of the subadditive ergodic theorem in [66]. For subadditiviy, note that

$$\begin{aligned} X_{0,n} &= -\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(ns,nt) \geqslant nr) \\ &\leqslant -\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(ms,mt) \geqslant mr) - \log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G((n-m)s,(n-m)t)\circ\theta_{ms,mt} \geqslant (n-m)r) \\ &= X_{0,m} + X_{m,n}. \end{aligned}$$

For $k \in \mathbb{N}$, by the ergodicity assumptions on μ, the sequence $(X_{k,k+n})_{n\in\mathbb{N}}$ has the same distribution as $(X_{0,n})_{n\in\mathbb{N}}$ and the sequence $(X_{(n-1)k,nk})_{n\in\mathbb{N}}$ is ergodic. Moreover, $X_{0,n} \geqslant 0$ and

$$\mathrm{E}\, X_{0,n} \leqslant \mathrm{E}\left[-\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(W(1,1) \geqslant nr)\right] = n \max\{r, 0\}\, \mathrm{E}[a+b] < \infty. \tag{2.5.54}$$

Hence, by the subadditive ergodic theorem, (2.5.52) holds μ-a.s. (and in expectation under μ) with

$$\mathbf{J}_{s,t}(r) = \lim_{n\to\infty} \frac{1}{n}\, \mathrm{E}\, X_{0,n} = \lim_{n\to\infty} -\frac{1}{n}\, \mathrm{E} \log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(ns,nt) \geqslant nr). \tag{2.5.55}$$

We record some properties of $\mathbf{J}_{s,t}(r)$ for $s,t \in \mathbb{N}$ and $r \in \mathbb{R}$. It is clear from (2.5.55) that $\mathbf{J}_{s,t}(r)$ is nonincreasing in (s,t) and nondecreasing in r. In addition, $\mathbf{J}_{s,t}(r) = 0$ for $r \leqslant 0$ as G is nonnegative, and $\mathbf{J}_{cs,ct}(cr) = c\,\mathbf{J}_{s,t}(r)$ for $c \in \mathbb{N}$. By (2.5.54), $\mathbf{J}_{s,t}(r) \leqslant r\, \mathrm{E}[a+b] < \infty$ for $r \geqslant 0$. Also, for $s_1, s_2, t_1, t_2 \in \mathbb{N}$ and $r_1, r_2 \in \mathbb{R}$, we have

$$\begin{aligned}\mathrm{E}\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(n(s_1+s_2), n(t_1+t_2)) \geqslant n(r_1+r_2)) \geqslant{}& \mathrm{E}\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(ns_1,nt_1) \geqslant nr_1) \\ &\cdot \mathrm{E}\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(ns_2,nt_2) \geqslant nr_2)\end{aligned}$$

for $n \in \mathbb{N}$, which gives $\mathbf{J}_{s_1+s_2,t_1+t_2}(r_1+r_2) \geqslant \mathbf{J}_{s_1,t_1}(r_1)+\mathbf{J}_{s_2,t_2}(r_2)$. Then, for $0 \leqslant r \leqslant r' \leqslant r+\frac{1}{n}$,

$$\begin{aligned}\mathbf{J}_{s,t}(r') - \mathbf{J}_{s,t}(r) &\leqslant \mathbf{J}_{s,t}(r+1/n) - \mathbf{J}_{s,t}(r) \\ &= \frac{1}{n+1}\,\mathbf{J}_{(n+1)s,(n+1)t}(nr+r+1+1/n) - \mathbf{J}_{s,t}(r) \\ &\leqslant \frac{\mathbf{J}_{ns,nt}(nr)}{n+1} - \mathbf{J}_{s,t}(r) + \frac{\mathbf{J}_{s,t}(r+2)}{n+1} \\ &= \frac{\mathbf{J}_{s,t}(r+2) - \mathbf{J}_{s,t}(r)}{n+1} \\ &\leqslant \frac{2r+2}{n}\, \mathrm{E}[a+b],\end{aligned} \tag{2.5.56}$$

which shows continuity of $\mathbf{J}_{s,t}(r)$ in r.

There exists a μ-a.s. event E on which (2.5.52) holds for all $s,t \in \mathbb{N}$ and $r \in \mathbb{Q}$. It follows from the monotonicity of $\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(ns,nt) \geqslant nr)$ in r and continuity of $\mathbf{J}_{s,t}$ that (2.5.52) holds for all $s,t \in \mathbb{N}$ and $r \in \mathbb{R}$ on E. From now on, let us work with $(\mathbf{a},\mathbf{b}) \in E$.

For $c > 0$, $\delta \in (0,1)$ and large enough $n \in \mathbb{N}$, we have

$$\begin{aligned}
-\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor ncs \rfloor, \lfloor nct \rfloor) \geqslant nr) &\leqslant -\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor cn \rfloor s, \lfloor cn \rfloor t) \geqslant \lfloor cn \rfloor r(1+\delta)) \\
-\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor ncs \rfloor, \lfloor nct \rfloor) \geqslant nr) &\geqslant -\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lceil cn \rceil s, \lceil cn \rceil t) \geqslant \lceil cn \rceil r(1-\delta)).
\end{aligned} \tag{2.5.57}$$

It follows from these inequalities and continuity of $\mathbf{J}_{s,t}$ that (2.5.52) holds on E with $\mathbf{J}_{cs,ct}(cr) = c\,\mathbf{J}_{s,t}(r)$. In particular, $\mathbf{J}_{s,t}(r)$ exists for rational $s, t > 0$. Moreover, by homogeneity, the properties of $\mathbf{J}_{s,t}(r)$ noted in preceding paragraph hold for rational $s, t > 0$ as well.

For $s, t, \delta > 0$, choose rational s', t' such that $\frac{s'}{1+\delta} < s \leqslant s'$ and $\frac{t'}{1+\delta} < t \leqslant t'$. Then

$$\begin{aligned}
-\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor ns \rfloor, \lfloor nt \rfloor) \geqslant nr) &\geqslant -\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor ns' \rfloor, \lfloor nt' \rfloor) \geqslant nr) \\
-\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor ns \rfloor, \lfloor nt \rfloor) \geqslant nr) &\leqslant -\log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor ns'/(1+\delta) \rfloor, \lfloor nt'/(1+\delta) \rfloor) \geqslant nr).
\end{aligned} \tag{2.5.58}$$

It follows that

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{1}{n} \log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor ns \rfloor, \lfloor nt \rfloor) \geqslant nr) \geqslant \mathbf{J}_{s',t'}(r) \\
&\limsup_{n\to\infty} -\frac{1}{n} \log \mathbf{P}_{\mathbf{a},\mathbf{b}}(G(\lfloor ns \rfloor, \lfloor nt \rfloor) \geqslant nr) \leqslant \mathbf{J}_{s'/(1+\delta),t'/(1+\delta)}(r) = \mathbf{J}_{s',t'}((1+\delta)r)/(1+\delta).
\end{aligned}$$

Using (2.5.56), we obtain

$$\frac{\mathbf{J}_{s',t'}((1+\delta)r)}{1+\delta} - \mathbf{J}_{s',t'}(r) \leqslant \mathbf{J}_{s',t'}((1+\delta)r) - \mathbf{J}_{s',t'}(r) \leqslant \frac{2r+2}{\lceil (r\delta)^{-1} \rceil} \mathrm{E}[a+b].$$

As $\delta \downarrow 0$, we have $s' \downarrow s$ and $t' \downarrow t$. Hence, we conclude that $\mathbf{J}_{s,t}(r)$ exists and equals the limit of $\mathbf{J}_{s',t'}(r)$, and also enjoys the properties of mentioned above. Finally, it follows from subadditivity and homogeneity that $\mathbf{J}$ is convex. $\square$

Proposition 2.5.15.

(a) μ-a.s., for any $s, t > 0$ and $\lambda \in \mathbb{R}$, there exists $\mathbf{L}_{s,t}(\lambda) \in [-\infty, \infty]$ such that,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}_{\mathbf{a},\mathbf{b}}[e^{\lambda G(\lfloor ns \rfloor, \lfloor nt \rfloor)}] = \mathbf{L}_{s,t}(\lambda) \tag{2.5.59}$$

(b) For any $s, t > 0$ and $\lambda \in \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{\lambda G(\lfloor ns \rfloor, \lfloor nt \rfloor)}] = \mathbb{L}_{s,t}(\lambda) \tag{2.5.60}$$

(c) $\mathbf{L}_{s,t}(\lambda)$ and $\mathbb{L}_{s,t}(\lambda)$ are nondecreasing and convex in λ.

(d) $\lambda\, \mathbf{L}_{s,t}(\lambda)$ and $\lambda\, \mathbb{L}_{s,t}(\lambda)$ are nondecreasing, homogeneous and concave in (s, t).

Proof. Fix $\lambda \in \mathbb{R}$ and $s, t \in \mathbb{N}$. Define

$$X_{m,n} = -\lambda \log \mathbf{E}_{\tau_{ms}(\mathbf{a}),\tau_{mt}(\mathbf{b})} \left[e^{\lambda G((n-m)s,(n-m)t)} \right]$$

for integers $0 \leqslant m < n$. Then $\{X_{m,n} : 0 \leqslant m < n\}$ are nonpositive and subadditive, and the conditions of the subadditive ergodic theorem are in place to claim the existence of $\mathbf{L}_{s,t}(\lambda) \in [-\infty, \infty]$ such that (2.5.59) holds μ-a.s.

For $\lambda \in \mathbb{R}$, $s, t \in \mathbb{N}$ and $c > 0$, we have

$$-\lambda \log \mathbf{E}_{\mathbf{a},\mathbf{b}} \left[e^{\lambda G(\lceil nc \rceil s, \lceil nc \rceil t)} \right] \leqslant -\lambda \log \mathbf{E}_{\mathbf{a},\mathbf{b}} \left[e^{\lambda G(\lfloor ncs \rfloor, \lfloor nct \rfloor)} \right] \leqslant -\lambda \log \mathbf{E}_{\mathbf{a},\mathbf{b}} \left[e^{\lambda G(\lfloor nc \rfloor s, \lfloor nc \rfloor t)} \right]$$

Also, for $\lambda \in \mathbb{R}$, $s, s', t, t', \delta > 0$ such that s', t' are rational, $\frac{s'}{1+\delta} < s \leqslant s'$ and $\frac{t'}{1+\delta} < t \leqslant t'$,

$$-\lambda \log \mathbf{E}_{\mathbf{a},\mathbf{b}} \left[e^{\lambda G(\lfloor ns' \rfloor, \lfloor nt' \rfloor)} \right] \leqslant -\lambda \log \mathbf{E}_{\mathbf{a},\mathbf{b}} \left[e^{\lambda G(\lfloor ns \rfloor, \lfloor nt \rfloor)} \right] \leqslant -\lambda \log \mathbf{E}_{\mathbf{a},\mathbf{b}} \left[e^{\lambda G(\lfloor \frac{ns'}{1+\delta} \rfloor, \lfloor \frac{nt'}{1+\delta} \rfloor)} \right].$$

Using these inequalities as in the preceding proof, we obtain (2.5.59) for all $s, t > 0$ μ-a.s. and the claimed properties of the function $(s, t) \mapsto \lambda\, \mathbf{L}_{s,t}(\lambda)$.

Now fix $s, t > 0$. Note that $\mathbf{L}_{s,t}(\lambda)$ is nondecreasing in λ. Let $\lambda_0 = \sup_{\lambda \in \mathbb{R}} \{\mathbf{L}_{s,t}(\lambda) < \infty\}$. For $\lambda_1, \lambda_2 \in \mathbb{R}$ and $c_1, c_2 \in (0, 1)$ with $c_1 + c_2 = 1$, by Hölder's inequality,

$$\log \mathbf{E}_{\mathbf{a},\mathbf{b}} \left[e^{(c_1\lambda_1 + c_2\lambda_2) G(\lfloor ns \rfloor, \lfloor nt \rfloor)} \right] \leqslant c_1 \log \mathbf{E}_{\mathbf{a},\mathbf{b}} \left[e^{\lambda_1 G(\lfloor ns \rfloor, \lfloor nt \rfloor)} \right] + c_2 \log \mathbf{E}_{\mathbf{a},\mathbf{b}} \left[e^{\lambda_2 G(\lfloor ns \rfloor, \lfloor nt \rfloor)} \right],$$

which implies that $\mathbf{L}_{s,t}(c_1\lambda_1 + c_2\lambda_2) \leqslant c_1\, \mathbf{L}_{s,t}(\lambda_1) + c_2\, \mathbf{L}_{s,t}(\lambda_2)$. Hence, $\mathbf{L}_{s,t}(\lambda)$ is continuous in λ on $(-\infty, \lambda_0)$. Using this and the monotonicity of last-passage times, we deduce that (2.5.59) holds for all $s, t > 0$ and $\lambda \in \mathbb{R}$ μ-a.s. $\square$

2.5.8 Lyapunov exponents for the stationary model

We close this section with the proof of Theorem 2.2.17

Proof of Theorem 2.2.32. We begin with the coupling

$$\begin{aligned}\hat{G}(\lfloor ns \rfloor, \lfloor nt \rfloor) = &\max_{1 \leqslant k \leqslant \lfloor ns \rfloor} \Big\{ G(\lfloor ns \rfloor - k + 1, \lfloor nt \rfloor) \circ \theta_{k-1,0} + \hat{G}(k, 0) \Big\} \\ &\vee \max_{1 \leqslant k \leqslant \lfloor nt \rfloor} \Big\{ G(\lfloor ns \rfloor, \lfloor nt \rfloor - k + 1) \circ \theta_{0,k-1} + \hat{G}(0, k) \Big\}.\end{aligned}$$

Arguing with lim sup and lim inf and coarse graining as above, this leads to the variational problem

$$\mathbf{L}^z_{s,t}(\lambda) = \max_{0 \leqslant r \leqslant s} \left\{ \mathbf{L}_{s-r,t}(\lambda) + r\, \mathrm{E}\left[\log \frac{a+z}{a+z-\lambda}\right] \right\} \vee \max_{0 \leqslant u \leqslant t} \left\{ \mathbf{L}_{s,t-u}(\lambda) + u\, \mathrm{E}\left[\log \frac{b-z}{b-z-\lambda}\right] \right\}.$$

Substituting in the variational expression for $\mathbf{L}_{s,t}(\lambda)$, this leads to

$$\begin{aligned}&\mathbf{L}^z_{s,t}(\lambda) = \\ &\max_{0 \leqslant r \leqslant s} \left\{ \min_{\theta \in [-\underline{\alpha}, \underline{\beta} - \lambda]} \left\{ (s-r)\, \mathrm{E}\left[\log \frac{a+\theta+\lambda}{a+\theta}\right] + t\, \mathrm{E}\left[\log \frac{b-\theta}{b-\theta-\lambda}\right] \right\} + r\, \mathrm{E}\left[\log \frac{a+z}{a+z-\lambda}\right] \right\} \\ &\vee \max_{0 \leqslant u \leqslant t} \left\{ \min_{\theta \in [-\underline{\alpha}, \underline{\beta} - \lambda]} \left\{ s\, \mathrm{E}\left[\log \frac{a+\theta+\lambda}{a+\theta}\right] + (t-u)\, \mathrm{E}\left[\log \frac{b-\theta}{b-\theta-\lambda}\right] \right\} + u\, \mathrm{E}\left[\log \frac{b-z}{b-z-\lambda}\right] \right\}.\end{aligned}$$

Applying a minimax theorem (for example [84]), we obtain

$$\begin{aligned}&\mathbf{L}^z_{s,t}(\lambda) = \\ &\min_{\theta \in [-\underline{\alpha}, \underline{\beta} - \lambda]} \left\{ s\, \mathrm{E}\left[\log \frac{a+\theta+\lambda}{a+\theta}\right] + t\, \mathrm{E}\left[\log \frac{b-\theta}{b-\theta-\lambda}\right] + \max_{0 \leqslant r \leqslant s} r\, \mathrm{E}\left[\log \frac{(a+z)}{(a+z-\lambda)} \frac{(a+\theta)}{(a+\theta+\lambda)}\right] \right\} \\ &\vee \min_{\theta \in [-\underline{\alpha}, \underline{\beta} - \lambda]} \left\{ s\, \mathrm{E}\left[\log \frac{a+\theta+\lambda}{a+\theta}\right] + t\, \mathrm{E}\left[\log \frac{b-\theta}{b-\theta-\lambda}\right] + \max_{0 \leqslant u \leqslant t} u\, \mathrm{E}\left[\log \frac{(b-z)}{(b-z-\lambda)} \frac{(b-\theta-\lambda)}{(b-\theta)}\right] \right\}.\end{aligned}$$

Write $(a+z-\lambda)(a+\theta+\lambda) = (a+z)(a+\theta) + \lambda(z-\theta-\lambda)$ to see that the inner maximum of the first term occurs at $r = s$ if $z - \lambda \leqslant \theta$ and $r = 0$ if $z - \lambda \geqslant \theta$. Similarly, $\theta \mapsto (1 - \lambda(b-\theta)^{-1})$ is a decreasing function, so the inner maximum of the second term occurs at $u = t$ for $\theta \leqslant z$ and at $u = 0$ for $\theta \geqslant z$. Breaking the first minimum over $[-\underline{\alpha}, \underline{\beta} - \lambda]$ into a minimum over

$[-\underset{\sim}{\alpha}, z-\lambda]$ and a minimum over $[z-\lambda, \underset{\sim}{\beta}]$ and the second into a minimum over $[-\underset{\sim}{\alpha}, z]$ and a minimum over $[z, \underset{\sim}{\beta}-\lambda]$, we obtain

$$\min_{\theta\in[-\underset{\sim}{\alpha},\underset{\sim}{\beta}-\lambda]} \left\{ s\,\mathrm{E}\left[\log\frac{a+\theta+\lambda}{a+\theta}\right] + t\,\mathrm{E}\left[\log\frac{b-\theta}{b-\theta-\lambda}\right] + \max_{0\leqslant r\leqslant s} r\,\mathrm{E}\left[\log\frac{(a+z)}{(a+z-\lambda)}\frac{(a+\theta)}{(a+\theta+\lambda)}\right]\right\} =$$
$$\left\{ s\,\mathrm{E}\left[\log\frac{a+z}{a+z-\lambda}\right] + t\,\mathrm{E}\left[\log\frac{b-z+\lambda}{b-z}\right]\right\} \wedge \min_{\theta\in[-\underset{\sim}{\alpha},z-\lambda]} \left\{ s\,\mathrm{E}\left[\log\frac{a+\theta+\lambda}{a+\theta}\right] + t\,\mathrm{E}\left[\log\frac{b-\theta}{b-\theta-\lambda}\right]\right\}$$

and similarly, for the remaining term we have

$$\min_{\theta\in[-\underset{\sim}{\alpha},\underset{\sim}{\beta}-\lambda]} \left\{ s\,\mathrm{E}\left[\log\frac{a+\theta+\lambda}{a+\theta}\right] + t\,\mathrm{E}\left[\log\frac{b-\theta}{b-\theta-\lambda}\right] + \max_{0\leqslant u\leqslant t} u\,\mathrm{E}\left[\log\frac{(b-z)}{(b-z-\lambda)}\frac{(b-\theta-\lambda)}{(b-\theta)}\right]\right\} =$$
$$\left\{ s\,\mathrm{E}\left[\log\frac{a+z+\lambda}{a+z}\right] + t\,\mathrm{E}\left[\log\frac{b-z}{b-z-\lambda}\right]\right\} \wedge \min_{\theta\in[z,\underset{\sim}{\beta}-\lambda]} \left\{ s\,\mathrm{E}\left[\log\frac{a+\theta+\lambda}{a+\theta}\right] + t\,\mathrm{E}\left[\log\frac{b-\theta}{b-\theta-\lambda}\right]\right\}$$

The function $\theta \mapsto s\,\mathrm{E}\left[\log\frac{a+\theta+\lambda}{a+\theta}\right] + t\,\mathrm{E}\left[\log\frac{b-\theta}{b-\theta-\lambda}\right]$ is strictly convex with a unique minimizer. Note that the first terms in each of these minima are the values of this function evaluated at $\theta = z-\lambda$ and $\theta = z$. The result follow from strict convexity by considering whether the minimizer lies in $[-\underset{\sim}{\alpha}, z], [z, z-\lambda]$, or $[z-\lambda, \underset{\sim}{\beta}-\lambda]$. $\square$

Chapter 3

Particle representations for a class of stochastic partial differential equations

3.1 Introduction

This chapter studies particle representations for a class of stochastic partial differential equations. The idea behind the approach taken here originates in the study of the McKean-Vlasov problem. A simple (deterministic) version of such a problem is to consider the following partial differential equation written in weak form

$$\langle \varphi, V(t) \rangle = \langle \varphi, V(0) \rangle + \int_0^t \langle \mathcal{L}(V(s))\varphi, V(s) \rangle ds, \tag{3.1.1}$$

where $\varphi \in C_c^\infty(\mathbb{R}^d)$, $\langle \cdot, \cdot \rangle$ denotes the pairing of a function with a measure, and $\mathcal{L}$ is the second order differential operator given by

$$\mathcal{L}(\nu)\varphi(x) = \frac{1}{2} \sum_{i,j=1}^{d} a_{i,j}(\nu, x)\partial_i \partial_j \varphi(x) + \sum_{i=1}^{d} c_i(\nu, x)\partial_i(x).$$

One approach to constructing a solution to this non-linear PDE which will work under certain regularity assumptions on the coefficients $a_{i,j}$ and c_i is to construct a collection of exchangeable diffusions $\{X_i(\cdot)\}$ which satisfy

$$X_i(t) = X_i(0) + \int_0^t c(V(s), X_i(s))ds + \int_0^t \sigma(V(s), X_i(s))dB_i(s), \tag{3.1.2}$$

where $\{B_i\}$ is a family of i.i.d. standard Brownian motions in $\mathbb{R}^d$, $[\sigma^t\sigma]_{i,j} = a_{i,j}$, and $V(t)$ is the de Finetti measure of the exchangeable sequence $\{X_i(t)\}$:

$$V(t) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(t)}. \tag{3.1.3}$$

Assume that such a system of diffusions $\{X_i\}$ has been constructed and $V(0)$ is the given initial condition. Then for sufficiently regular functions φ we may apply Itô's lemma to obtain

$$\varphi(X_i(t)) = \varphi(X_i(0)) + \int_0^t \mathcal{L}(V(s))\varphi(X_i(s))ds + M_{\varphi,i}(t) \tag{3.1.4}$$

where $\{M_{\varphi,i}\}_i$ are mean zero orthogonal martingales. By taking averages of both sides of (3.1.3), one can see that if $V(t)$ is given by (3.1.3) where the $X_i(t)$ solve (3.1.2), then $V(t)$ will be a solution to (3.1.1).

The mathematical study of systems of diffusions of this type began with the seminal work of McKean [68], though there are many approaches to this problem. See for example [36, 57, 73]. Such systems appear in a wide variety of applications, ranging from physics to economics. See for example [19] and the references above. One advantage of the particle framework is that it provides a model of microscopically interacting processes which aggregate to the solution to the stochastic partial differential equation; it is often of interest in applications to observe this phenomenon.

Our specific interest is in a stochastic perturbation of the construction described above. The general approach taken here originally appears in [62] and the treatment presented below can be viewed as an extension of the results of [63]. We introduce a perturbation of the de Finetti measure of the sequence $\{X_i(t)\}$ by introducing family of processes of weights $\{A_i(t)\}$ with the property that the family $\{(X_i, A_i)\}$ is exchangeable. The measure which will serve as a solution to the class of stochastic partial differential equations under consideration will be defined by

$$V(t) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} A_i(t)\delta_{X_i(t)}. \tag{3.1.5}$$

Assuming that the $E[|A_i(t)|] < \infty$, then de Finetti's theorem (or the ergodic theorem) implies that this defines a random (signed) measure. Note that introducing weights allows for a second source of non-linearity. In [63], which studied stochastic partial differential equations on all of $\mathbb{R}^d$, the authors take advantage of both of these sources. For technical reasons, in the problems we consider (on domains $D \subset \mathbb{R}^d$) we will need to restrict to the case in which $\mathcal{L}(\nu) := \mathcal{L}$ does not depend on ν. To avoid introducing unnecessary notation, we will restrict to this case in the remainder of this introduction as well.

To provide some motivation for the results below, we sketch the construction of a simple case of the type of stochastic partial differential equation studied in [63]. Suppose that we have an exchangeable sequence of pairs of weights and particles (X_i, A_i), where the diffusions $\{X_i\}$ are i.i.d. and satisfy

$$X_i(t) = X_i(0) + \int_0^t c(X_i(s))ds + \int_0^t \sigma(X_i(s))dB_i(s), \tag{3.1.6}$$

where c, σ, σ^{-1} are continuous and bounded, and the weights A_i satisfy

$$A_i(t) = A_i(0) + \int_0^t G(V(s), X_i(s))A_i(s)ds + \int_{\mathbb{R}^d \times (0,t]} \rho(X_i(s) - u)A_i(s)W(du \times ds). \tag{3.1.7}$$

Here, we assume that G is sufficiently regular, ρ is a $C_c^\infty(\mathbb{R}^d)$ mollifier,

$$V(t) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n A_i(t)\delta_{X_i(t)}, \tag{3.1.8}$$

the B_i are i.i.d. standard Brownian motions, and W is a space-time white noise in the sense of Walsh [90] which is independent of $\{B_i\}$. Set

$$M_{\varphi,i}(t) = \varphi(X_i(t)) - \varphi(X_i(0)) - \int_0^t \mathcal{L}\,\varphi(X_i(s))ds,$$

and as before the family $\{M_{\varphi,i}\}_i$ are orthogonal martingales. Applying Itô's lemma,

$$A_i(t)\varphi(X_i(t)) = A_i(0)\varphi(X_i(0)) + \int_0^t \left(\mathcal{L}\,\varphi(X_i(s)) + \varphi(X_i(s))G(V(s), X_i(s))\right) A_i(s)ds$$

$$+ \int_{\mathbb{R}^d \times (0,t]} \varphi(X_i(s))\rho(X_i(s) - u)A_i(s)W(du \times ds) + \int_0^t A_i(s)dM_{\varphi,i}(s).$$

Averaging, we see that the measure $V(t)$ will solve

$$\begin{aligned}\langle \varphi, V(t)\rangle = \langle \varphi, V(0)\rangle + \int_0^t \langle \mathcal{L}\,\varphi(\cdot) + \varphi(\cdot)G(V(s),\cdot), V(s)\rangle ds \\ + \int_{\mathbb{R}^d \times (0,t]} \langle \varphi(\cdot)\rho(\cdot - u), V(s)\rangle W(du \times ds).\end{aligned} \tag{3.1.9}$$

Under mild conditions, one can show existence and uniqueness of a measure-valued process $V(\cdot)$ given by (3.1.8) with A_i given by (3.1.7) and X_i given by (3.1.6) [63, Theorems 2.1,2.2]. As a consequence, one may view $V(\cdot)$ and the collection $\{(X_i, A_i)\}$ interchangeably. This is the sense in which the particle system $\{(X_i, A_i)\}$ forms a *representation* of a solution $V(t)$ to the weak form stochastic partial differential equation (3.1.9). Similarly, if one can show uniqueness for solutions of the weak form stochastic partial differential equation (3.1.9), then the particle system is a representation of *the* solution to this equation.

For equations of the type considered above, and again under mild conditions, one can show that $V(t)$ is absolutely continuous with respect to the Lebesgue measure [63, Theorem 3.5]. More specifically, if $V(0)(dx) = V(0,x)dx$ for some $V(0,\cdot) \in L^2(\mathbb{R}^d)$, then $V(t)(dx) = V(t,x)dx$ where $V(t,\cdot) \in L^2(\mathbb{R}^d)$. If we let $\mathcal{L}^*$ denote the $L^2(\mathbb{R}^d)$ adjoint of $\mathcal{L}$, the measure $V(t) := V(t,x)dx$ can then be viewed as a weak solution to the stochastic partial differential equation

$$\partial_t V(t,x) = \mathcal{L}^*\, V(t,x) + V(t,x)G(V(t,y)dy, x) + \int_{\mathbb{R}^d} \rho(x-u)V(t,x)W(dt \times du).$$

With [63] having constructed particle representations for a wide class of stochastic partial differential equations on $\mathbb{R}^d$ it is natural to wonder whether a similar construction is possible on domains $D \subset \mathbb{R}^d$ with boundary conditions on ∂D. This is achieved for a certain class of models with Dirichlet boundary conditions and additive white noise forcing (as opposed to the multiplicative forcing discussed in the previous example) in [28].

To see how this works, we consider the example which served as one of the main motivations of the construction in [28]: the stochastic Allen-Cahn equation on a smooth open domain D:

$$\partial_t u = \Delta u + u - u^3 + \xi,$$

$$u(0,x) = h(x), \qquad x \in D$$

$$u(t,x) = g(x), \qquad x \in \partial D, t > 0$$

where ξ is space-colored time-white noise. In order to impose Dirichlet boundary conditions on a particle solution to a stochastic partial differential equation similar to those considered above, one begins with a family of i.i.d. reflecting diffusions $\{X_i\}$ on a domain D and weights similar to those above, with one major difference—whenever a particle hits the boundary, the associated weight is assigned a value coming from the boundary condition. Let D be a smooth bounded domain, let g be a continuous function on ∂D, and let $\{X_i\}$ be a family of i.i.d. stationary normally reflecting Brownian motions run at speed $2t$. Introduce the notation $\tau_i(t) = 0 \vee \sup\{s \leqslant t : X_i(s) \in \partial D\}$ and suppose that $A_i(t)$ is a solution to

$$A_i(t) = g(X_i(\tau_i(t)))1_{\{\tau_i(t)>0\}} + \int_{\tau_i(t)}^t \left(1 - V(s, X_i(s))^2\right) A_i(s)ds \tag{3.1.10}$$
$$+ \int_{\mathbb{R}^d \times (\tau_i(t),t]} \rho(X_i(s) - u)A_i(s)W(du \times ds),$$

where $(t,x) \mapsto V(t,x)$ is a process of measurable versions of the densities with respect to the uniform distribution on D of the measure given by

$$V(t) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n A_i(t)\delta_{X_i(t)}, \tag{3.1.11}$$

and ρ is again a $C_c^\infty(\mathbb{R}^d)$ mollifier. To simplify the notation slightly, we have also imposed the condition that $\langle \varphi, V(0)\rangle = 0$ here. This corresponds to $h(x) = 0$.

One can show that if $\varphi \in C_c^\infty(D)$ (that is, φ is a smooth function which is compactly supported on the interior of D), then

$$\varphi(X_i(t))A_i(t) = \int_0^t \left(\Delta\varphi(X_i(s)) + (1 - V(s, X_i(s))^2)\varphi(X_i(s))\right) A_i(s)ds \tag{3.1.12}$$

$$+\int_{\mathbb{R}^d \times (0,t]} \varphi(X_i(s))\rho(X_i(s)-u)A_i(s)W(du \times ds).$$

Averaging, we have

$$\begin{aligned}\langle \varphi, V(t)\rangle = &\int_0^t \langle \Delta\varphi(\cdot) + (1-V(s,\cdot)^2)\varphi(\cdot), V(s)\rangle ds \\ &+ \int_{\mathbb{R}^d \times (0,t]} \langle \varphi(\cdot)\rho(\cdot - u), V(s)\rangle W(du \times ds).\end{aligned} \tag{3.1.13}$$

Moreover, one can show that the boundary condition is satisfied in the sense that for any $\overline{g} \in C(\overline{D})$ with $\overline{g}|_{\partial D} = g$, and for all t

$$\lim_{\epsilon \to 0} \mu\left(\{x \in D : d(x, \partial D) < \epsilon\}\right)^{-1} \int_{\{x \in D: d(x,\partial D)<\epsilon\}} |V(t,x) - \overline{g}(x)| dx = 0$$

in the L^1 sense [28, Proposition 2.15], where μ denotes the Lebesgue measure.

The restriction to $\varphi \in C_c^\infty(D)$ in (3.1.13) is to account for the fact that $A_i(t)$ is not a semi-martingale. By only considering φ which vanish in a neighborhood of ∂D, we avoid needing to address what happens to the stochastic integral near times at which $X_i(t) \in \partial D$. These times are the only obstruction to applying the usual semi-martingale integral results to expressions like the one in (3.1.12). This problem can be addressed directly for a wider class of test functions. The identity (3.1.13) can also be extended to $\varphi \in C_b^2(\mathbb{R}_+ \times \overline{D})$ with $\varphi(t,\cdot)|_{\partial D} = 0$ to obtain a weak form of the stochastic partial differential equation with a boundary term to account for the Dirichlet boundary condition:

$$\begin{aligned}\langle \varphi(t,\cdot), V(t)\rangle = &\int_0^t \langle \Delta\varphi(s,\cdot) + \partial_t\varphi(s,\cdot) + (1-V(s,\cdot)^2)\varphi(\cdot), V(s)\rangle ds \\ &+ \int_{\mathbb{R}^d \times (0,t]} \langle \varphi(\cdot)\rho(\cdot - u), V(s)\rangle W(du \times ds) \\ &+ \int_0^t \int_{\partial D} g(x)\nabla\varphi(s,x)\cdot\eta(x)\beta(dx)ds\end{aligned} \tag{3.1.14}$$

where η is the inward unit normal on ∂D and β is a measure which is proportional to the surface measure. It can be shown that a unique measure valued process $V(t)$ given by (3.1.11) with $A_i(t)$ given by (3.1.10) and moreover that (up to a moment condition) this measure valued process is the unique solution to the weak-form stochastic partial differential equation in (3.1.14) [28, Theorem 3.1].

More generally, the results of [28] give weak solutions to a class of stochastic partial differential equations with additive space-colored time-white noise of the form

$$\partial_t V(t,x) = \mathcal{L}^* V(t,x) + V(t,x)G(V(t,x),x,t) + b(x) + \int_{\mathbb{R}^d} \rho(x-u)W(dt \times du), \tag{3.1.15}$$
$$V(t,0) = h(x), \qquad x \in D,$$
$$V(t,x) = g(x), \qquad x \in \partial D, t > 0,$$

where D is a sufficiently regular open domain, $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$ in an appropriate L^2 space, $g \in C_b(\partial D)$ $\|h\|_\infty, \|b\|_\infty < \infty$ and the non-linearity $G(v,x,t)$ satisfies $G(v,x,t) \leqslant K$ and $|G(v,x,t)| \leqslant K(1+|v|^2)$ for some $K > 0$ and all $v \in \mathbb{R}$, $x \in \overline{D}$ and $t \geqslant 0$. The stochastic Allen-Cahn equation discussed above corresponds to the choices $G(v,x,t) = 1 - v^2$, and $h(x) = b(x) = 0$.

Some of the techniques of [28] appear to break down for equations with multiplicative noise. In particular, the proof of existence and uniqueness of the particle system [28, Theorem 2.2] relies on moment estimates which cannot be expected to hold if the noise is multiplicative. The goal of what follows is to use an alternate approach to extend the class of models for which particle representations can be derived to include certain equations with multiplicative noise.

3.2 Results

The goal of this chapter is to prove existence of a particle representation for weak solutions to a class of non-linear stochastic partial differential equations of the form

$$\begin{aligned}
\partial_t V(t,x) &= \mathcal{L}^* V(t,x) + G(V(t,y)\pi(dy), x, t)V(t,x) + b(x) \\
&\quad + \int_{\mathbb{R}^d} \rho(V(t,y)\pi(dy), x-u)V(t,x)W(dt \times du), \\
V(0,x) &= h(x), \qquad x \in D, \\
V(t,x) &= g(x), \qquad x \in \partial D, t > 0,
\end{aligned} \tag{3.2.1}$$

where π is the stationary distribution of a reflecting diffusion on $\overline{D}$, $\mathcal{L}^*$ is the adoint with respect to π of the generator of that reflecting diffusion, and ρ and G are sufficiently regular (but possibly non-linear) functions of the measure $V(t,y)\pi(dy)$ along with space and time.

As mentioned in the introduction, technical difficulties arise in applying the techniques of [28] to the class of problems we consider. We take a different approach to proving strong existence here. The outline of the argument is to show weak existence of a particle representation and then pathwise uniqueness of solutions which are jointly compatible with driving noise (the reflecting diffusions and the white noise). We begin by outlining the precise assumptions that will be used in what follows and by defining compatibility.

3.2.1 Assumptions, notation, and definitions

Before stating the assumptions precisely, we note that these results are not stated (or proven) in the greatest generality possible. In particular, it should be possible to extend the spatial coloring of the white noise to the same level of generality as in [63].

Assumptions and notation

Let $D = \{x : \phi(x) > 0\}$ for some $\phi \in C_b^2(\mathbb{R}^d)$ with $|\nabla\phi(x)| \geqslant 1$ for all $x \in \partial D$. Denote by $\mathcal{M}_+(D)$ the collection of finite non-negative Borel measures on $\overline{D}$. For a Borel measurable function ϕ on $\overline{D}$ and $\mu \in \mathcal{M}_+(D)$, we use the notation

$$\langle \phi, \mu \rangle = \int_{\overline{D}} \phi d\mu.$$

Denote by $\mathrm{Lip}(D)$ the Banach space of Lipschitz continuous functions on $\overline{D}$ equipped with the norm $\|\cdot\|_\infty + |\cdot|_L$ where $|\cdot|_L$ denotes the Lipschitz seminorm. We denote by $\mathrm{Lip}_1(D)$, the collection of functions $\{\phi \in \mathrm{Lip}(D) : \|\phi\|_\infty + |\phi|_L \leqslant 1\}$. Equip $\mathcal{M}_+(D)$ with the distance

$$d(\mu, \nu) = \sup_{\phi \in \mathrm{Lip}_1(D)} \langle \phi, \mu - \nu \rangle.$$

$d(\mu, \nu)$ generates the weak topology on $\mathcal{M}_+(D)$ [13, Theorem 8.3.2]. Let μ be a sigma finite positive measure on $\mathbb{R}^d$. We consider bounded Borel measurable functions $g : \partial D \mapsto \mathbb{R}_+$, $h : D \mapsto \mathbb{R}_+$, $b : \overline{D} \mapsto \mathbb{R}_+$, $G : \mathcal{M}_+(D) \times \overline{D} \times \mathbb{R}_+ \mapsto \mathbb{R}$, $\rho : \mathcal{M}_+(D) \times \mathbb{R}^d \to \mathbb{R}$. We will assume that $g \in C_b(\partial D)$ and that ρ and G are uniformly Lipschitz continuous. We will also require that there is a common compact set $\mathcal{K} \subset \mathbb{R}^d$ with $\mathrm{supp}\{\rho(\nu, \cdot)\} \subset \mathcal{K}$ for all $\nu \in \mathcal{M}_+(D)$. Notationally, we will let $K > 0$ be such that

1. For all $\nu \in \mathcal{M}_+(D)$,

$$\|g\|_\infty + \|h\|_\infty + \|b\|_\infty + \|G\|_\infty + \|\rho\|_\infty + \int_{\mathbb{R}^d} |\rho(\nu, u)|^2 \mu(du) \leqslant K$$

2. For all $\nu_1, \nu_2 \in \mathcal{M}_+(D)$, $x_1, x_2 \in \overline{D}$,

$$\begin{aligned} |G(\nu_1, x_1, s) - G(\nu_2, x_2, s)|^2 + |\rho(\nu_1, x_1) - \rho(\nu_2, x_2)|^2 & \\ + \int_{\mathbb{R}^d} |\rho(\nu_1, x_1 - u) - \rho(\nu_2, x_2 - u)|^2 \mu(du) & \\ \leqslant K(d(\nu_1, \nu_2)^2 + |x_1 - x_2|^2). & \end{aligned}$$

Let $\{X_i\}$ be a family of i.i.d. stationary reflecting diffusions in $\overline{D}$ with stationary distribution π solving the Skorokhod equation

$$X_i(t) = X_i(0) + \int_0^t c(X_i(s))ds + \int_0^t \sigma(X_i(s))dB_i(s) + \int_0^t \eta(X_i(s))dL_i(s), \tag{3.2.2}$$

where $\sup_{x\in\partial D} \eta(x)\cdot\nabla\phi(x) < 0$, and L_i is a local time for X_i on ∂D. We will assume that $c_i(\cdot), \sigma_{i,j}(\cdot)$ are Hölder continuous functions on $\overline{D}$, $\|c\|_\infty, \|\sigma\|_\infty < \infty$ and that the diffusion matrix σ is uniformly elliptic. Call $a_{i,j} := [\sigma^t\sigma]_{i,j}$ and

$$\mathcal{L} := \frac{1}{2}\sum_{i,j=1}^d a_{i,j}(x)\partial_i\partial_j + \sum_{i=1}^d c_i(x)\partial_i.$$

Under these hypotheses, the (sub-)martingale problem for X_i will be well-posed [86]. Denote the generator of this process by $\mathbb{A}$. We additionally have that the collection $\mathcal{D}(\mathbb{A}) = \{\varphi \in C_b^2(\overline{D}) : \nabla\varphi\cdot\eta|_{\partial D} = 0\}$ forms a core for $\mathbb{A}$ [33, Theorem 8.1.5]; that is, $\mathbb{A}$ is the closure of $\mathcal{L}$ defined on $\mathcal{D}(\mathbb{A})$. As a technical assumption, we require that if $X_i(0)$ is distributed according to π for each i and if the family $\{X_i\}$ is i.i.d., then $P(\exists i \neq j, i, j \in \mathbb{N}$ and $t > 0 : X_i(t), X_j(t) \in \partial D) = 0$.

We view the family $\{X_i\}$ as a random variable taking values in the complete separable metric space $C_{\overline{D}^\infty}[0,\infty)$ equipped with the Borel sigma algebra. For X_i given by (3.2.2), we define $\tau_i(t) = \inf\{s \leqslant t : X_i(s) \in \partial D\} \vee 0$.

Let W be a white noise (in the sense of Walsh) on $\mathbb{R}^d\times\mathbb{R}_+$ with respect to $\mu\otimes\lambda$ where λ is the Lebesgue measure on $\mathbb{R}_+$. We assume that W is independent of $\{X_i\}$. Let $\{e_n\}_{n=1}^\infty$ be an orthonormal basis for the separable Hilbert space $(L^2(\mathbb{R}^d,\mu), \langle\cdot,\cdot\rangle_L)$ and let $\mathcal{H} := \{x \in L^2(\mathbb{R}^d,\mu) : \|x\|_{\mathcal{H}}^2 = \sum_{n=1}^\infty n^2\langle x, e_n\rangle_L\} < \infty$. Define $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ by polarization and note that $\{n^{-1}e_n\}_{n=1}^\infty$ is an orthonormal basis for $\mathcal{H}$. It follows that the inclusion $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}) \hookrightarrow (L^2(\mathbb{R}^d,\mu), \langle\cdot,\cdot\rangle_L)$ is Hilbert-Schmidt. We denote by $\mathcal{H}^{-1}$ the continuous dual of $\mathcal{H}$. One can check that there exists a version of W with the property that pointwise the map $T \mapsto (x \in \mathcal{H} \mapsto \int_0^T\int_{\mathbb{R}^d} x(u)W(du\times dt)) \in C_{\mathcal{H}^{-1}}[0,\infty)$. Without loss of generality, we work with this version of W and will at times abuse notation and think of W as the map $T \mapsto (x \in \mathcal{H} \mapsto \int_0^T \int_{\mathbb{R}^d} x(u) W(du \times dt))$.

Definition of compatibility

Let $\mathcal{S}_1 = D_{\mathcal{M}_+}[0,\infty)$ and $\mathcal{S}_2 := C_{\mathcal{H}^{-1}}[0,\infty) \times C_{\overline{D}^\infty}[0,\infty)$. Then $\mathcal{S} := \mathcal{S}_1 \times \mathcal{S}_2$ is a complete separable metric space, which we equip with its Borel sigma algebra. Our definition of compatibility between inputs and outputs will follow the definition of temporal compatibility in [61].

Let $\{\mathcal{F}_t^{W,\{X_i\}}\}_{t\geqslant 0}$ denote the augmentation of the filtration $\sigma(W(C \times [0,s]), X_i(s) : F \in \mathcal{B}(D), C \in \mathcal{B}(\mathbb{R}^d), \mu(C) < \infty, 0 \leqslant s \leqslant t, i \in \mathbb{N})$. For $U \in D_{\mathcal{M}_+}[0,\infty)$, let $\mathcal{F}_t^{U,W,\{X_i\}}$ denote the augmentation of $\sigma(U(s)(F), W(C \times [0,s]), X_i(s) : F \in \mathcal{B}(D), C \in \mathcal{B}(\mathbb{R}^d), \mu(C) < \infty, 0 \leqslant s \leqslant t, i \in \mathbb{N})$. Similarly, let $\{\mathcal{F}_t^{U,V,W,\{X_i\}}\}_{t\geqslant 0}$ denote the augmentation of $\sigma(U(s)(F), V(s)(F), W(C \times [0,s]), X_i(s) : F \in \mathcal{B}(D), C \in \mathcal{B}(\mathbb{R}^d), \mu(C) < \infty, 0 \leqslant s \leqslant t, i \in \mathbb{N})$.

Definition 3.2.1. *A positive measure valued process $U(\cdot) \in D_{\mathcal{M}_+}[0,\infty)$ is* compatible *with $(W, \{X_i\})$ if for all $t > 0$ and all $h \in B_b(S_2)$, $E[h(W, \{X_i\}) | \mathcal{F}_t^{U,W,\{X_i\}}] = E[h(W, \{X_i\}) | \mathcal{F}_t^{W,\{X_i\}}]$*

Definition 3.2.2. *A pair of positive measure valued processes $U(\cdot) \in D_{\mathcal{M}_+}[0,\infty)$ and $V(\cdot) \in D_{\mathcal{M}_+}[0,\infty)$ are* jointly compatible *with $(W, \{X_i\})$ if for all $t > 0$ and all $h \in B_b(S_2)$,*
$E[h(W, \{X_i\}) | \mathcal{F}_t^{U,V,W,\{X_i\}}] = E[h(W, \{X_i\}) | \mathcal{F}_t^{W,\{X_i\}}]$.

For convenience, in what follows we will refer to processes as "compatible" or "jointly compatible" without reference to $(W, \{X_i\})$. In particular, if U is adapted to $\{\mathcal{F}_t^{W,\{X_i\}}\}_{t\geqslant 0}$, then U is compatible. We remark that compatibility or joint compatibility is sufficient to ensure that W remains white noise and the semi-martingale decomposition of X_i remains unchanged in the filtrations $\mathcal{F}_t^{U,W,\{X_i\}}$ or $\mathcal{F}_t^{U,V,W,\{X_i\}}$.

Definition 3.2.3. *A positive measure valued process $U(\cdot) \in D_{\mathcal{M}_+}[0,\infty)$ is* consistent *if it is compatible and the family $\{(W, X_i, U)\}_{i=1}^\infty$ is exchangeable.*

Definition 3.2.4. *A pair of positive measure valued process $U(\cdot) \in D_{\mathcal{M}_+}[0,\infty)$ and $V(\cdot) \in D_{\mathcal{M}_+}[0,\infty)$ are* jointly consistent *if they are jointly compatible and the family $\{(W, X_i, U, V)\}_{i=1}^{\infty}$ is exchangeable.*

Remark 3.2.5. Note that the additional requirement of exchangeability corresponds to a family of constraints in the language of [61].

3.2.2 Statement of results

For consistent U, we set

$$\Psi^U_{t_1,t_2} = \exp\bigg\{ \int_{t_1}^{t_2} G(U(s), X_i(s), s)ds - \frac{1}{2}\int_{\mathbb{R}^d} \rho(X_i(s) - u)^2 \mu(du) \\ + \int_{\mathbb{R}^d} \rho(U(s-), X_i(s), u) W(du \times ds)\bigg\}.$$

With this notation, if $A_i^U(t)$ is given by

$$A_i^U(t) = \left[g(X_i(\tau_i(t)))1_{\{\tau_i(t)>0\}} + h(X_i(0))1_{\{\tau_i(t)=0\}}\right] \Psi^U_{\tau_i(t),t} + \int_{\tau_i(t)}^t b(X_i(s))\Psi^U_{s,t} ds$$

then $A_i^U(t)$ solves

$$A_i^U(t) = g(X_i(\tau_i(t))1_{\{\tau_i(t)>0\}} + h(X_i(0))1_{\{\tau_i(t)=0\}} + \int_{\tau_i(t)}^t G(U(s), X_i(s), s) A_i^U(s) ds \tag{3.2.3}$$
$$+ \int_{\tau_i(t)}^t b(X_i(s))ds + \int_{\mathbb{U}\times(\tau_i(t),t]} \rho(U(s-), X_i(s) - u) A_i^U(s-) W(du \times ds).$$

For consistent U it will be convenient to define a reference process by

$$\Gamma_i^U(t) = \exp\left\{2 \sup_{0\leqslant s\leqslant t} \left| \int_{\mathbb{R}^d \times (0,s]} \rho(U(s-), X_i(s) - u) W(du \times ds)\right|\right\}$$

With this notation, we have the bound

$$A_i^U(t) \leqslant (\|g\|_\infty \vee \|h\|_\infty + t\|b\|_\infty)\, e^{Kt} \Gamma_i^U(t).$$

A key step in many of the proof of the results that follow is the observation that under the assumptions in Subsection 3.2.1 it is possible to prove uniform (conditional) moment bounds:

Lemma 3.2.6. *Suppose that U is consistent. Then*

$$E\left[\sup_{0\leqslant t\leqslant T} A_i^U(t)^p\right] \leqslant C_{T,p} := 4(1+T)Ke^{KT(1+8p^2)}.$$

If in addition U is adapted to the augmentation of the filtration $\{\mathcal{F}_t^W \vee \sigma(X_i)\}_{t\geqslant 0}$, then

$$E\left[\sup_{0\leqslant t\leqslant T} A_i^U(t)^p|\sigma(X_i)\right] \leqslant C_{T,p}.$$

For consistent U, $\{A_i^U\}$ given by (3.2.3), and $\{X_i\}$ given by (3.2.2), we note that the collection $\{(X_i, A_i^U)\}$ is exchangeable. By de Finetti's theorem (see for example [53, Theorem 9.16]), we may then define a measure $\Xi^U \in D_{\overline{D}\times\mathbb{R}_+}[0,\infty)$ via the almost sure limit

$$\Xi^U = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \delta_{X_i, A_i^U}.$$

Working on a full measure subset of the set on which this limit holds, we may define a new consistent process $\Phi U \in D_{\mathcal{M}_+(D)}[0,\infty)$ via

$$\Phi U(t) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n A_i^U(t)\delta_{X_i(t)}.$$

See also [62, Section 10, Subsection 11.3] for a similar construction. One can see directly that $\Phi U(t)$ is absolutely continuous with respect to π. Let $\Phi U(t,x)$ be a Borel measurable version of the density of $\Phi U(t)$ with respect to π; i.e., $\Phi U(t,x)\pi(dx) := \Phi U(t)(dx)$. We also note that if U is $\{\mathcal{F}_t^W\}_{t\geqslant 0}$ adapted, then ΦU will be $\{\mathcal{F}_t^W\}_{t\geqslant 0}$ adapted.

It will be useful to know that if U is consistent, then ΦU will be a continuous process of positive measures. Our proof of this depends on the technical assumption above that there are almost surely no times t for which $X_i(t) \in \partial D$ and $X_j(t) \in \partial D$, where $i \neq j$.

Lemma 3.2.7. *Suppose that U is consistent. Then $\Phi U \in C_{\mathcal{M}_+(D)}[0,\infty)$.*

ΦU will typically be a weak solution solution to a certain (usually *linear*) stochastic partial differential equation with multiplicative noise and the same boundary conditions as in (3.2.1). We begin by observing that the measure valued process ΦU solves a weak form stochastic partial differential equation on the interior of D:

Proposition 3.2.8. *Suppose that U is a consistent process in $C_{\mathcal{M}_+(D)}[0,\infty)$. Then for $\varphi \in C_c^\infty(D)$,*

$$\langle \varphi, \Phi U(t)\rangle = \langle \varphi, \Phi U(0)\rangle + \int_0^t \langle \mathcal{L}\,\varphi, \Phi U(s)\rangle ds + \int_0^t \langle \varphi(\cdot) G(U(s),\cdot), \Phi U(s)\rangle ds \qquad (3.2.4)$$
$$+\, t\langle \varphi b, \pi\rangle + \int_{\mathbb{R}^d \times [0,t]} \langle \varphi(\cdot)\rho(U(s),\cdot,u), \Phi U(s)\rangle W(du \times ds).$$

We will present two ways in which ΦU satisfies the boundary conditions. These are essentially the same as in [28] and the proofs are more or less identical. The first depends on regularity of the time-reversal of the driving diffusions. For each t and $s \leqslant t$, define the time reversal of X_i by $X^*_{i,t}(s) = X_i(t-s)$. We will often suppress the subscript and define $X^*_{i,t}(s) := X^*_i(s)$. Since X_i is stationary, the time reversal X^*_i is a Markov process whose generator $\mathbb{A}^*$ satisfies

$$\int_D g\,\mathbb{A}\,f d\pi = \int_D f\,\mathbb{A}^*\,g d\pi, \quad f,g \in \mathcal{D}(\mathbb{A}).$$

We introduce the notation $\sigma_i = \inf\{s : X^*_i(s) \in \partial D\}$, and note that if we reverse time starting from t, then $\sigma_i = t - \tau_i(t)$. The first sense in which the boundary condition is satisfied depends on the following condition.

Condition 3.2.9. *The boundary ∂D is regular for X^*_i in the sense that for each $\delta > 0$ and $x \in \partial D$,*

$$\lim_{y \in D \to x} P\left(\sigma_i > \delta | X^*_i(0) = y\right) = 0, \qquad (3.2.5)$$

and

$$\lim_{y \in D \to x} E\left[|X^*_i(\sigma_i) - x| \wedge 1 | X^*_i(0) = y\right] = 0. \qquad (3.2.6)$$

Remark 3.2.10. In practice this condition can be difficult to check unless $\mathbb{A} = \mathbb{A}^*$, as regularity of the time reversed process involves some knowledge about regularity of a density for π with respect to the Lebesgue measure.

Remark 3.2.11. If D is bounded and X_i is normally reflecting Brownian motion run at speed $2t$, then $\mathbb{A} = \mathbb{A}^*$, $\mathcal{L} = \Delta$, π is proportional to the Lebesgue measure, and Condition 3.2.9 holds.

Proposition 3.2.12. *Suppose that Condition 3.2.9 is satisfied and suppose that U is $\{\mathcal{F}_t^W\}_{t\geqslant 0}$ adapted. Let $\bar{g}$ be any bounded continuous function on $\overline{D}$ with $\bar{g}|_{\partial D} = g$. Then for any compact $K \subset \partial D$ and $t > 0$,*

$$\lim_{\epsilon\to 0} \frac{\int_{\partial_\epsilon(K)} |\Phi U(t,x) - \bar{g}(x)|\pi(dx)}{\pi(\partial_\epsilon(K))} = 0 \tag{3.2.7}$$

in $L^1(P)$.

As was the case in [28], we can extend the weak formulation of the stochastic partial differential equation (3.2.4) to include a boundary term. To do this, we need to introduce the boundary measure which is associated to the local time L_i. It is shown in [86] that under our assumptions, $E[L_i(t)] < \infty$. By stationarity, for $t \in \mathbb{R}_+$, the process $X_i(t+\cdot)$ has the same distribution as $X_i(\cdot)$. It follows that for $\varphi \in C_b(\partial D)$, $E\left[\int_s^t \varphi(X_i(r))dL_i(r)\right] = E\left[\int_0^{t-s} \varphi(X_i(r))dL_i(r)\right]$. By non-negativity and linearity in t and φ, the Riesz representation theorem implies that there exists a finite positive Radon measure β on ∂D with the property that for $\varphi \in C_b(\mathbb{R}_+ \times \partial D)$,

$$E\left[\int_0^t \varphi(X_i(s),s)dL_i(s)\right] = \int_0^t \int_{\partial D} \varphi(x,s)\beta(dx)ds.$$

Remark 3.2.13. If D is bounded and X_i is normally reflecting standard Brownian motion, β is proportional to the surface measure.

Proposition 3.2.14. *Suppose that U is a consistent process in $C_{\mathcal{M}_+(D)}[0,\infty)$. Then for $\varphi \in C_c^2(\mathbb{R}_+ \times \overline{D})$ with $\varphi(t,\cdot)|_{\partial D} = 0$ for all $t \geqslant 0$,*

$$\begin{aligned}\langle \varphi(\cdot,t), \Phi U(t)\rangle = \langle \varphi(\cdot,0), \Phi U(0)\rangle &+ \int_0^t \langle \mathcal{L}\varphi(\cdot,s) + \partial_t\varphi(\cdot,s), \Phi U(s)\rangle ds \\ &+ \int_0^t \langle \varphi(\cdot,s)G(U(s),\cdot), \Phi U(s)\rangle ds + \int_0^t \langle \varphi(\cdot,s)b(\cdot),\pi\rangle ds\end{aligned} \tag{3.2.8}$$

$$+\int_{\mathbb{R}^d \times [0,t]} \langle \varphi(\cdot,s)\rho(U(s),\cdot,u), \Phi U(s)\rangle W(du \times ds)$$
$$+\int_0^t \int_{\partial D} g(x)\eta(x) \cdot \nabla\varphi(x,s)\beta(dx)ds.$$

Having shown that ΦU satisfies (3.2.4) and (3.2.8) and that under Condition 3.2.9 and the hypothesis that U is $\mathcal{F}_t^W$ adapted (3.2.7) holds, we now look to show existence of a measure valued process satisfying $U = \Phi U$. For such a process, (3.2.4) is a weak formulation of a stochastic partial differential equation of the form in (3.2.1) on the interior of D. The final results of this chapter are Theorems 3.2.16 and 3.2.15, which show pathwise uniqueness for jointly consistent solutions and weak existence respectively. These combine to prove strong existence of a measure valued process which represents a solution to this non-linear stochastic partial differential equation.

Theorem 3.2.15. *Suppose that U and V are jointly consistent and that there is $T > 0$ so that for all $t \leqslant T$, $\Phi U(t) = U(t)$ and $\Phi V(t) = V(t)$. Then almost surely for all $t \leqslant T$, $U(t) = V(t)$.*

Theorem 3.2.16. *There exists a consistent measure valued process $U(\cdot) \in C_{\mathcal{M}_+(D)}[0,\infty)$ which satisfies $U(t) = \Phi U(t)$ for all t.*

The results of [61] imply that the previous results combine to prove the desired strong existence of the particle representation:

Theorem 3.2.17. *There exists a Borel measurable function $F : C_{\mathcal{H}^{-1}}[0,\infty) \mapsto D_{\mathcal{M}_+(D)}[0,\infty)$ with the property that $F(W)(\cdot)$ is $\{\mathcal{F}_t^W\}_{t\geqslant 0}$ adapted and $F(W)(\cdot) = \Phi F(W)(\cdot)$.*

3.3 Proofs

We begin with some preliminary results, starting with an observation about the structure of ΦU.

Lemma 3.3.1. *Suppose that U is consistent. Then for fixed t and $\varphi \in B_b(\overline{D})$ almost surely*

$$\langle \varphi, \Phi U(t) \rangle = E[A_1^U(t)\varphi(X_1(t))|\sigma(W) \vee \sigma(U)].$$

Moreover, if $U(t)$ is $\{\mathcal{F}_t^W\}_{t \geqslant 0}$ adapted, then

$$\Phi U(t, X_i(t)) = E\left[A_i^U(t)|W, X_i(t)\right].$$

Proof. Let $\mathcal{I}$ denote the shift invariant sigma algebra for the stationary sequence $\{(X_i, W, U)\}_i$. By the ergodic theorem,

$$\begin{aligned}\langle \varphi, \Phi U(t) \rangle = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n A_i^U(t)\varphi(X_i(t)) &= E[A_1^U(t)\varphi(X_1(t))|\mathcal{I}] \\ &= E[A_1^U(t)\varphi(X_1(t))|\sigma(W) \vee \sigma(U)].\end{aligned}$$

The last equality follows from the fact that the sequence $\{X_i\}_i$ is i.i.d.. Let $F : C_{\mathcal{H}^{-1}}[0, \infty) \mapsto \mathbb{R}$ be bounded and Borel measurable. Using the assumption that U is $\{\mathcal{F}_t^W\}_{t \geqslant 0}$ adapted and the first part of the result, we have

$$E\left[A_i^U \varphi(X_i(t))F(W)\right] = E\left[\int_D \Phi U(t, x)\varphi(x)\pi(dx)F(W)\right] = E\left[\Phi U(t, X_i(t))\varphi(X_i(t))F(W)\right].$$

In the last line, we use independence of W and X_i. $\square$

3.3.1 Moment estimates

Recall the definition of the weights A_i^U. Define

$$H(t) = \int_{\mathbb{R}^d \times [0,t]} \rho(U(s-), X_i(s) - u) W(du \times ds),$$

$$[H]_t = \int_0^t \int_{\mathbb{R}^d} \rho^2(U(s), X_i(s) - u)\mu(du)ds,$$

and let $\{\mathcal{H}_t\}_{t\geqslant 0}$ be the right continuous completion of $\{\mathcal{F}_t^W \vee \sigma(\{X_i\})\}_{t\geqslant 0}$ with respect to P. Note that W is $\mathcal{H}_t$ white noise and if $U \in D_{\mathcal{M}_+}[0,\infty)$ is adapted to $\{\mathcal{F}_t^W \vee \sigma(\{X_i\})\}_{t\geqslant 0}$, then $H(t)$ is a continuous $\mathcal{H}_t$ martingale. Define $\gamma_i(t) = \inf\{s > 0 : \int_0^s \int_{\mathbb{R}^d} \rho^2(U(r), X_i(r) - u)\mu(du)dr > t\}$. The next result is a slight modification of the Dubins-Schwarz theorem and we follow the proof of that result. See for example [78, Theorems V.1.6,V.1.10].

Lemma 3.3.2. *Suppose that U is consistent. Then for each i there exists a standard Brownian motion B with $B(0) = 0$ and*

$$B\left(\int_0^t \int_{\mathbb{R}^d} \rho^2(U(s), X_i(s) - u)\mu(du)ds\right) = H_i(t)$$

and moreover

$$\sup_{0\leqslant t\leqslant T} |H_i(t)| \leqslant \sup_{0\leqslant s\leqslant Kt} |B(s)|.$$

If in addition U is adapted to $\{\mathcal{H}_t\}_{t\geqslant 0}$, then this Brownian motion may be taken to be independent of $\sigma(\{X_i\})$.

Proof. Suppose first that $[H]_\infty = \infty$ a.s.. If U is consistent with W, then $B(t) = H_i(\gamma_i(t))$ is a continuous martingale with respect to $\{\mathcal{F}_{\gamma_i(t)}^{U,W,\{X_i\}}\}_{t\geqslant 0}$. If U is $\{\mathcal{H}_t\}_{t\geqslant 0}$ adapted, then $B_i(t)$ is a martingale with respect to $\{\mathcal{H}_{\gamma_i(t)}\}_{t\geqslant 0}$. Moreover, $[B]_t = t$. The inequality follows from $\int_{\mathbb{R}^d} |\rho(\nu, x - u)|^2\mu(du) < K$.

By continuity of H_i and the definition of $\gamma_i(0)$, $B(0) = 0$. We note that $\sigma(\{X_i\}) \subset \mathcal{H}_{\gamma(0)}$ and consequently if U is adapted to $\{\mathcal{H}_t\}$ then the Brownian motion B_i is independent of $\sigma(\{X_i\})$.

If $P([H]_\infty < \infty) > 0$, then take $j \neq i$ and let $B(t) = H_i(\gamma_i(t))1_{\{t<[H]_\infty\}} + (\lim_{t\to\infty} H_i(t) + B_j(t - [H]_\infty))1_{\{t\geqslant[H]_\infty\}}$. See [78, Proposition IV.1.26] for a proof that the limit exists. Then the arguments above continue to hold with minor modifications to the filtration that do not change the conclusions. □

Proof of Lemma 3.2.6. It follows from Lemma 3.3.2 that there is a Brownian motion B such that

$$\sup_{0\leqslant t\leqslant T} A_i^U(t)^p \leqslant (\|g\|_\infty \vee \|h\|_\infty + T\|b\|_\infty)e^{KT}\exp\left\{2p \sup_{0\leqslant s\leqslant KT}|B(s)|\right\}.$$

The two parts of the lemma then follow from taking expectations or conditional expectations respectively. □

3.3.2 Continuity and a weak formulation of the stochastic partial differential equation

We now turn to Lemma 3.2.7, which says that for consistent U, $\Phi U \in C_{\mathcal{M}_+(D)}[0,\infty)$.

Proof of Lemma 3.2.7. Recall the notation $d(\nu,\mu)$, which denotes is the Fortet-Mourier distance on $\mathcal{M}_+(D)$, and introduce

$$\Phi^{(n)}U(t) = \frac{1}{n}\sum_{i=1}^n A_i^U(t)\delta_{X_i(t)}, \tag{3.3.1}$$

so that $\Phi^{(n)}U(\cdot) \to \Phi U(\cdot)$ almost surely in $D_{\mathcal{M}_+(D)}[0,\infty)$. A necessary and sufficient condition for $\Phi U \in C_{\mathcal{M}_+(D)}[0,\infty)$ is that the random variable

$$\int_0^\infty e^{-u}[\sup_{0\leqslant r\leqslant u} d(\Phi U^n(r), \Phi U^{(n)}(r-)) \wedge 1]du$$

converges to zero in distribution [33, Theorem 3.10.2]. In particular, by bounded convergence it suffices to show that for fixed $t > 0$, $\sup_{0\leqslant r\leqslant t} d(\Phi U^n(r), \Phi U^{(n)}(r-)) \wedge 1 \to 0$ in probability. We have

$$d(\Phi^{(n)}U(r), \Phi^{(n)}(r-)) = \sup_{\varphi\in \mathrm{Lip}_1(D)} \frac{1}{n}\sum_{i=1}^n \left(A_i^U(r) - A_I^U(r-)\right)\varphi(X_i(r)) \leqslant \frac{1}{n}\sum_{i=1}^n |A_i^U(r) - A_i^U(r-)|.$$

Next we use the technical assumption in Subsection 3.2.1 that $P(\exists i, j \in \mathbb{N}, i \neq j, t > 0 : X_i(t), X_j(t) \in \partial D) = 0$ and the observation that $A_i(\cdot)$ is continuous off of the set $\{t : X_i(t) \in \partial D\}$ to obtain:

$$\sup_{0 \leqslant r \leqslant t} d(\Phi^{(n)} U(r), \Phi^{(n)}(r-)) \leqslant \frac{1}{n} \max_{i \leqslant n} \sup_{0 \leqslant r \leqslant t} |A_i^U(r) - A_i^U(r-)| \leqslant \frac{2}{n} \max_{i \leqslant n} \sup_{0 \leqslant r \leqslant t} A_i^U(r).$$

Applying a union bound, Markov's inequality, and Lemma 3.2.6, we see that

$$P\left(\frac{2}{n} \max_{i \leqslant n} \sup_{0 \leqslant r \leqslant t} A_i^U(r) > \epsilon\right) \to 0$$

as $n \to \infty$. □

Now, we turn to the proof of Proposition 3.2.8, which gives a weak formulation of the stochastic partial differential equation with test functions $\varphi \in C_c^\infty(D)$. We begin with a type of decomposition which will be of use in some later proofs as well. Fix a consistent process $U \in D_{\mathcal{M}_+(D)}[0, \infty)$. Define $Z_i^U(t)$ by

$$\begin{aligned} Z_i^U(t) = &\int_0^t G(U(s), X_i(s), s) A_i^U(s) ds + \int_0^t b(X_i(s)) ds \\ &+ \int_{\mathbb{R}^d \times (0,t]} \rho(U(s-), X_i(s) - u) A_i^U(s-) W(du \times ds). \end{aligned} \tag{3.3.2}$$

Then we have

$$A_i^U(t) = Z_i^U(t) - Z_i^U(\tau_i(t)) + g(X_i(\tau_i(t))) 1_{\{\tau_i(t) > 0\}} + h(X_i(0)) 1_{\{\tau_i(t) = 0\}}. \tag{3.3.3}$$

With this decomposition in hand, we can now address the weak formulation of the stochastic partial differential equation on the interior of D.

Proof of Proposition 3.2.8. Recall that the assumption that U is compatible implies that the semi-martingale decomposition of X_i in $\{\mathcal{F}_t^{U,W,\{X_i\}}\}_{t \geqslant 0}$ is given by (3.2.2) and note that $Z_i^U(t)$ is a semi-martingale in this filtration and that the covariation of $Z_i(\cdot)$ and $\varphi(X_i(\cdot))$ is zero. Take $\varphi \in C_c^2(D)$ and a partition $\{t_j\}$ of $[0, T]$.

$$A_i(T) \varphi(X_i(T)) = A_i(0) \varphi(X_i(0)) + \sum_j A_i(t_j)(\varphi(X_i(t_{j+1})) - \varphi(X_i(t_j)))$$

$$+\sum_j \varphi(X_i(t_{j+1}))(A_i(t_{j+1}) - A_i(t_j)).$$

As the mesh of the partition tends to zero, the first sum converges to an ordinary Ito integral because $A_i^U(\cdot-)$ is $\{\mathcal{F}_t^{U,W,\{X_i\}}\}_{t\geqslant 0}$ predictable. Note that the last three terms (3.3.3) are piecewise constant off of the set $\{t : X_i(t) \in \partial D\}$. Indeed, for $t < t'$, $\tau_i(t) \neq \tau_i(t')$ if and only if there is $s \in (t, t']$ with $X_i(s) \in \partial D$. By local uniform continuity of $t \mapsto \varphi(X_i(t))$ and the fact that $\varphi(x)$ is compactly supported in the interior of D, we see that there is a (random) $\delta > 0$ so that for any $t \leqslant T$ with $X_i(t) \in \partial D$ and any s with $|s - t| < \delta$, $\varphi(X_i(s)) = 0$. Using the fact that the covariation of $\varphi(X_i(\cdot))$ and $Z_i(\cdot)$ is zero, we may use the usual semi-martingale integral results to obtain a version of Ito's lemma for $A_i^U(t)\varphi(X_i(t))$:

$$\begin{aligned} A_i^U(t)\varphi(X_i(t)) = {} & A_i^U(0)\varphi(X_i(0)) + \int_0^t \mathcal{L}\,\varphi(X_i(s))A_i^U(s)ds \\ & + \int_0^t G(U(s), X_i(s), s)\varphi(X_i(s))A_i^U(s) + \varphi(X_i(s))b(X_i(s))ds \\ & + \int_{\mathbb{R}^d\times(0,t]} \rho(U(s), X_i(s) - u)\varphi(X_i(s))A_i^U(s-)W(du \times ds) \\ & + \int_0^t A_i^U(s-)dM_{\varphi,i}(s) \end{aligned} \tag{3.3.4}$$

where $\{M_{\varphi,i}\}_i$ is a family of orthogonal martingales given by

$$M_{\varphi,i}(t) = \varphi(X_i(t)) - \varphi(X_i(0)) - \int_0^t \mathcal{L}\,\varphi(X_i(s))ds.$$

Recall the notation $\Phi U^{(n)}$ from (3.3.1) and let $\pi^{(n)}(t) = n^{-1}\sum_{i=1}^n \delta_{X_i(t)}$. Averaging (3.3.4) gives

$$\begin{aligned} \langle\varphi, \Phi U^{(n)}(t)\rangle = {} & \langle\varphi, \Phi U^{(n)}(0)\rangle + \int_0^t \langle\mathcal{L}\,\varphi, \Phi U^{(n)}(s)ds \\ & + \int_0^t \langle G(U(s), \cdot, s)\varphi(\cdot), \Phi U^{(n)}(s) + \langle\varphi b, \pi^{(n)}(s)\rangle ds \\ & + \int_{\mathbb{R}^d\times(0,t]} \langle\rho(U(s), \cdot - u)\varphi(\cdot), \Phi^{(n)}U(s-)\rangle W(du \times ds) \\ & + \frac{1}{n}\sum_{i=1}^n \int_0^t A_i^U(s-)dM_{\varphi,i}(s) \end{aligned} \tag{3.3.5}$$

Note that for each i, $Y_i(\cdot)\int_0^\cdot A_i^U(s-)dM_{\varphi,i}(s)$ is a mean zero $\{\mathcal{F}_t^{U,W,\{X_i\}}\}_{t\geqslant 0}$ martingale and that the family $\{Y_i\}_i$ is orthogonal. Doob's inequality shows that $n^{-1}\sum_{i\leqslant n} Y_i(\cdot) \to 0$ locally uniformly in probability. Using the fact that the limit $\Phi^{(n)}U(\cdot) \to \Phi U(\cdot)$ occurs in $D_{\mathcal{M}_+(D)}[0,\infty)$, one can show that the averages of all of the terms except the second to last converges to the corresponding term in (3.2.4). To show local uniform convergence of the white noise term, it suffices to show that

$$\int_0^t E\left[\int_{\mathbb{R}^d} \langle \varphi(\cdot)\rho(U(s),\cdot - u), \Phi U^{(n)}(s) - \Phi U(s)\rangle^2 \mu(du)\right] ds \to 0.$$

For each fixed $s \geqslant 0$ and $u \in \mathbb{R}^d$, the function $x \mapsto \varphi(x)\rho(U(s), x-u)$ is continuous and compactly supported and therefore bounded. It follows from the fact that $\Phi U^{(n)} \to \Phi U$ in $D_{\mathcal{M}_+(D)}[0,\infty)$ that the integrand tends to zero pointwise almost everywhere with respect to $ds \otimes \mu \otimes P$. By a hypothesis on ρ and using the fact that φ is compactly supported, we may restrict to a common compact subset of $\overline{D} \times \mathbb{R}^d$ and therefore it suffices to prove uniform integrability. We have

$$\begin{aligned}\langle \varphi(\cdot)\rho(U(s),\cdot - u), \Phi U^{(n)}(s) - \Phi U(s)\rangle^4 &\leqslant \langle |\varphi(\cdot)||\rho(U(s),\cdot - u)|, \Phi U^{(n)}(s) + \Phi U(s)\rangle^4 \\ &\leqslant K^4 \|\varphi\|_\infty^4 \langle 1, \Phi U^{(n)}(s) + \Phi U(s)\rangle^4.\end{aligned}$$

The moment bound from Lemma 3.2.6 completes the proof. □

3.3.3 Boundary condition

Averaged boundary condition

We now turn to the proofs of the boundary conditions, which were discussed previously in Subsection 3.2.2.

Proof of Proposition 3.2.12. We have

$$E\left[\pi(\partial_\epsilon(K))^{-1}\int_{\partial_\epsilon(K)} |\Phi U(t,x) - \bar{g}(x)|\pi(dx)\right] \leqslant E\left[|A_i^U(t) - \bar{g}(X_i(t))|\big|X_i(t) \in \partial_\epsilon(K)\right].$$

Recalling the definition of $A_i^U(t)$, we have

$$|A_i^U(t) - \bar{g}(X_i(t))| \leqslant |g(X_i(\tau_i(t))) - \bar{g}(X_i(t))| 1_{\{\tau_i(t)>0\}} + 2(K \vee \|\bar{g}\|_\infty) 1_{\{\tau_i(t)=0\}} \tag{3.3.6}$$
$$+ K(t - \tau_i(t)) + K \int_{\tau_i(t)}^t A_i^U(s) ds$$
$$+ \Big| \int_{\mathbb{R}^d \times (\tau_i(t), t]} \rho(U(s-), X_i(s) - u) A_i^U(s-) W(du \times ds) \Big|.$$

Next, observe that

$$E\left[|g(X_i(\tau_i(t))) - \bar{g}(X_i(t))| 1_{\{\tau_i(t)>0\}} \Big| X_i(t) \in \partial_\epsilon(K) \right]$$
$$= E\left[|g(X_i^*(\sigma_i)) - \bar{g}(X_i^*(0))| 1_{\{\sigma_i<t\}} \Big| X_i^*(0) \in \partial_\epsilon(K) \right]$$
$$\leqslant \sup_{x \in \partial_\epsilon K} E\left[|g(X_i^*(\sigma_i)) - \bar{g}(x)| 1_{\{\sigma_i<t\}} \Big| X_i^*(0) = x \right].$$

By compactness, there exists $x_0 \in K$ and $x_n \to x_0$ so that

$$\limsup_{\epsilon \to 0} \sup_{x \in \partial_\epsilon K} E\left[|g(X_i^*(\sigma_i)) - \bar{g}(x)| 1_{\{\sigma_i<t\}} \Big| X_i^*(0) = x \right]$$
$$= \lim_{n \to \infty} E\left[|g(X_i^*(\sigma_i)) - \bar{g}(x_n)| 1_{\{\sigma_i<t\}} \Big| X_i^*(0) = x_n \right].$$

Continuity of $\bar{g}$ and (3.2.6) imply that the limit is zero. A similar argument and (3.2.5) show that $P(\tau_i(t) = 0 | X_i(t) \in \partial_\epsilon K)$ tends to zero, so that the conditional expectations of the first two terms on the right of (3.3.6) go to zero. Note $t - \tau_i(t) \in [0, t]$ and further that

$$E\left[\int_{\tau_i(t)}^t A_i^U(s) ds \Big| X_i(t) \in \partial_\epsilon(K) \right] \leqslant$$
$$E\left[t - \tau_i(t) \Big| X_i(t) \in \partial_\epsilon(K) \right]^{\frac{1}{2}} E\left[\int_0^t A_i^U(s)^2 ds \Big| X_i(t) \in \partial_\epsilon(K) \right]^{\frac{1}{2}}.$$

As above, the first term tends to zero by hypothesis. The second term is uniformly bounded by Lemma 3.2.6. Notice that W remains white noise when conditioned on $\sigma(X_i)$. Recall that the L^2 norm dominates the L^1 and note that

$$E\left[\Big| \int_{\mathbb{R}^d \times (0,t]} 1_{(\tau_i(t), t]}(s) \rho(U(s-), X_i(s) - u) A_i^u(s-) W(du \times ds) \Big|^2 \Big| X_i(t) \in \partial_\epsilon(K) \right]$$

$$\leqslant E\left[\int_0^t \int_{\mathbb{R}^d} 1_{(\tau_i(t),t]}(s)\rho(U(s), X_i(s)-u)^2 A_i^U(s)^2\mu(du)ds \middle| X_i(t) \in \partial_\epsilon(K)\right]$$
$$\leqslant KE\left[\int_0^t 1_{(\tau_i(t),t]}(s)A_i^U(s)^2 ds \middle| X_i(t) \in \partial_\epsilon(K)\right].$$

Arguing as above, the last term tends to zero as $\epsilon \to 0$. □

Weak formulation with boundary

In order to prove the weak formulation in Proposition 3.2.14, we need some preliminary lemmas. The goal here is to show that for most values of t with $X_i(t) \in \partial D$, we also have $A_i(t) = g(X_i(t))$. Note that this is not *always* the case. The proof of the next result is essentially the same as the proof of the analogous properties for the zero set of one dimensional Brownian motion.

Lemma 3.3.3. *Almost surely, the set $\{t \geqslant 0 : X_i(t) \in \partial D\}$ is a closed set with no isolated points and the collection of left-isolated points of this set is countable.*

Proof. Under the assumptions in Subsection 3.2.1, X_i satisfies the strong Markov property and the boundary is regular for X_i [86, Theorems 2.4,5.8, Corollary 2.3]. Recall that ∂D is a closed set and X_i is continuous, so this set is closed. Set $\alpha_t^X = \inf\{s > t : X(s) \in \partial D\}$. By the strong Markov property, for each $q \in \mathbb{Q}_+$, α_q^X is not right-isolated. Fix $t_0 \in \{t \geqslant 0 : X_i(t) \in \partial D\}$ with $t_0 \neq \alpha_q^X$ for any $q \in \mathbb{Q}_+$. Take a sequence $q_n \in \mathbb{Q}_+$ with $q_n \uparrow t_0$. Then $q_n \leqslant \alpha_{q_n}^X < t_0$ and therefore t_0 is not left isolated. □

Lemma 3.3.4. *Almost surely,*

$$\int_{\{t:\tau_i(t)\neq t\}} dL_i(s) = \int_{\{t:\tau_i(t-)\neq t\}} dL_i(s) = 0.$$

Proof. Local time is a continuous measure supported on the set $\{t \geqslant 0 : X_i(t) \in \partial D\}$ and therefore assigns measure zero to the (countable) set of left isolated points of $\{t \geqslant 0 : X_i(t) \in \partial D\}$. If $t_0 \in \{t \geqslant 0 : X_i(t) \in \partial D\}$ is not left-isolated, then $t_0 = \tau_i(t_0) = \tau_i(t_0-)$. □

The previous results combine to prove the following lemma.

Lemma 3.3.5. *Almost surely, for dL_i almost every t, $A_i^U(t) = A_i^U(t-) = g(X_i(t))$ and therefore*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \int_0^t A_i^U(s-)\eta(X_i(s)) \cdot \nabla\varphi(X_i(s), s)dL_i(s) = \int_0^t \int_{\partial D} g(x)\eta(x) \cdot \nabla\varphi(x, s)\beta(dx)ds.$$

We now turn to the proof of Proposition 3.2.14. The structure of the argument is similar to that of Proposition 3.2.8. The bulk of the argument is essentially the proof of [28, Lemma 3.3].

Proof of Proposition 3.2.14. We proceed as in the proof of Proposition 3.2.8 above. Take φ as in the statement of the result and a partition $\{t_j\}$ of $[0, T]$. Summation by parts gives

$$\begin{aligned} A_i(T)\varphi(X_i(T), T) = A_i(0)\varphi(X_i(0), 0) + \sum_j A_i(t_j)(\varphi(X_i(t_{j+1}), t_{j+1}) - \varphi(X_i(t_j)), t_j) \\ + \sum_j \varphi(X_i(t_{j+1}), t_{j+1})(A_i(t_{j+1}) - A_i(t_j)). \end{aligned}$$

As in the proof of Proposition 3.2.8, the first sum converges to the usual Ito integral by standard results. Note that this convergence implies convergence of the last term, because this expression is an identity. We will again take advantage of the decomposition (3.3.3) to compute the limit of the last term. Once again, noting that the covariation of $\varphi(X_i(\cdot))$ and $Z_i(\cdot)$ is zero, the contribution of the term coming from $Z_i^U(t)$ follows from the usual semi-martingale integral results. We have

$$\begin{aligned} &\sum_j \varphi(X_i(t_{j+1}), t_{j+1})(Z_i^U(\tau_i(t_{j+1})) - Z_i^U(\tau_i(t_j)) \\ &= \sum_j \varphi(X_i(t_{j+1}), t_{j+1}) \int_{\tau_i(t_j)}^{\tau_i(t_{j+1})} G(U(s), X_i(s), s)A_i^U(s))ds \\ &+ \sum_j \varphi(X_i(t_{j+1}), t_{j+1}) \int_{\tau_i(t_j)}^{\tau_i(t_{j+1})} b(X_i(s))ds \end{aligned}$$

$$+\sum_j \varphi(X_i(t_{j+1}),t_{j+1})\int_{\mathbb{R}^d\times(\tau_i(t_j),\tau_i(t_{j+1})]}\rho(U(s-),X_i(s)-u)A_i^U(s-)W(du\times ds).$$

Recall that $\tau_i(t_{j+1})=\tau_i(t_j)$ unless there exists $t\in(\tau_i(t_j),\tau_i(t_{j+1})]$ with $X_i(t)\in\partial D$. For any such t, the hypothesis gives $\varphi(X_i(t),t)=0$. Note that for each j, the absolute value of the first integral can be bounded by $KT\sup_{0\leqslant t\leqslant T}A_i^U(t)$ and similar bounds can be derived for the other integrals. It follows that the limit of these terms as the mesh tends to zero is zero.

The term involving h is piecewise constant with a single jump discontinuity, at which time $X_i(t)\in\partial D$, so a similar argument shows that the contribution from that term tends to zero. It remains to show that the contribution from the term $g(X_i(\tau_i(t)))1_{\tau_i(t)>0}$ tends to zero. To simplify the notation, we drop the indicator function of the set $\{\tau_i(t)>0\}$. Accounting for this change is an argument similar to the argument showing convergence of the term involving h, but note that this does change the initial condition in the expression we compute below. With summation by parts, we have

$$\begin{aligned}&\sum_j \varphi(X_i(t_j),t_j)(g(X_i(\tau_i(t_{j+1})))-g(X_i(\tau_i(t_j))))\\&=\varphi(X_i(T),T)g(X_i(\tau_i(T))-\varphi(X_i(0),0)g(X_i(0))\\&-\sum_j g(X_i(\tau_i(t_j)))(\varphi(X_i(t_{j+1})-\varphi(X_i(t_j)).\end{aligned}$$

As the mesh tends to zero, the last term converges to the usual semi-martingale integral of $g(X_i(\tau_i(\cdot-)))$ with respect to $\varphi(X_i(\cdot))$. Introduce the notation $\gamma_i(t)=\inf\{s\geqslant t: X_i(s)\in\partial D\}$ and define

$$U_n(t)=\sum_{k=0}^{\infty} g\left(X_i\left(\gamma_i\left(\frac{k}{n}\right)\right)\right)1_{[\gamma_i(\frac{k}{n}),\gamma_i(\frac{k+1}{n}))}(t).$$

Observe that $U_n(t)\to g(X_i(\tau_i(t)))$ pointwise. Moreover, we have

$$\begin{aligned}&\int_0^t (U_n(s-)-g(X_i(\tau_i(s-)))d\varphi(X_i(t),t)\\&=\int_0^t (U_n(s-)-g(X_i(\tau_i(s-)))\nabla\varphi(X_i(s),s)^t\sigma(X_i(s))dB_i(s)\end{aligned}$$

$$+ \int_0^t (U_n(s-) - g(X_i(\tau_i(s)))(\mathcal{L} + \partial_t)\varphi(X_i(s), s)ds$$
$$+ \int_0^t (U_n(s-) - g(X_i(\tau_i(s-)))\nabla\varphi(X_i(s), s) \cdot \eta(X_i(s))dL_i(s).$$

By Doob's inequality and bounded convergence, these tend to zero termwise locally uniformly in probability. Noting that $\varphi(X_i(\gamma_i(\cdot)), \cdot) \equiv 0$, we have

$$\int_0^t g\left(X_i\left(\gamma_i\left(\frac{k}{n}\right)\right)\right) 1_{(\gamma_i(\frac{k}{n}),\gamma_i(\frac{k+1}{n})]}(s-)d\varphi(X_i(s), s)$$
$$= \begin{cases} g\left(X_i\left(\gamma_i\left(\frac{k}{n}\right)\right)\right)\varphi(X_i(t), t) & \text{if } t \in (\gamma_i(\frac{k}{n}), \gamma_i(\frac{k+1}{n})] \\ 0 & \text{otherwise} \end{cases}.$$

and therefore, we have

$$\int_0^t U_n(s-)d\varphi(X_i(s), s) = U_n(t)\varphi(X_i(t), t) - g(X_i(0))\varphi)(X_i(0), 0).$$

Combining these results with Lemma 3.3.5, we obtain a version of Ito's lemma for $A_i^U(t)\varphi(X_i(s), s)$:

$$A_i^U(t)\varphi(X_i(t), t) = h(X_i(0))\varphi(X_i(0), 0) + \int_0^t A_i^U(s)\varphi(X_i(s), s)(G(U(s), X_i(s), s) + b(X_i(s)))ds$$
$$+ \int_0^t (\mathcal{L} + \partial_t)\varphi(X_i(s), s)A_i^U(s)ds$$
$$+ \int_{\mathbb{R}^d \times (0,t]} A_i^U(s-)\rho(U(s), X_i(s), s)\varphi(X_i(s), s)W(du \times ds)$$
$$+ \int_0^t g(X_i(s))\nabla\varphi(X_i(s), s) \cdot \eta(X_i(s))dL_i(s)$$
$$+ \int_0^t A_i^U(s-)\nabla\varphi(X_i(s))\sigma(X_i(s))dB_i(s).$$

Averaging and arguing as in the proof of Proposition 3.2.8 completes the proof. □

3.3.4 Uniqueness of a particle fixed point

Let U and V be jointly consistent. Take $\varphi \in \mathrm{Lip}_1(D)$ and observe that for all $t \in [0, T]$ we have

$$|\langle\varphi, \Phi U(t) - \Phi V(t)\rangle| = \lim_{n\to\infty} \frac{1}{n} \left|\sum_{i=1}^n \left(A_i^U(t) - A_i^V(t)\right)\varphi(X_i(t))\right| \leqslant \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \left|A_i^U(t) - A_i^V(t)\right|.$$

It then follows that

$$d(\Phi U(t), \Phi V(t)) \leqslant \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left|A_i^U(t) - A_i^V(t)\right|.$$

We begin by proving pathwise uniqueness of jointly consistent functions, which was already stated as Lemma 3.2.15.

Proof of Lemma 3.2.15. Define

$$\eta_m = \inf\left\{t > 0 : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (\Gamma_i^U(t) \vee \Gamma_i^V(t))^2 > m\right\}.$$

Compatibility ensures that η_m is a stopping time and W is white noise in the filtration $\{\mathcal{F}_t^{U,V,W,\{X_i\}}\}_{t\geqslant 0}$. Using the inequality $|e^x - e^y| \leqslant e^x \vee e^y|x-y|$, we see that there is a deterministic constant C (which may change line to line) such that for $t \leqslant T$

$$\begin{aligned}|A_i^U(t) - A_i^V(t)| \leqslant C\Gamma_i^U(t) \vee \Gamma_i^V(t) \times \Bigg(\int_0^t |G(U(s), X_i(s), s) - G(V(s), X_i(s), s)|\, ds \\ + \int_0^t \int_{\mathbb{R}^d} |\rho^2(U(s), X_i(s) - u) - \rho^2(V(s), X_i(s) - u)|\mu(du) ds \\ + \sup_{0\leqslant s\leqslant t} \left|\int_{\mathbb{R}^d \times (s,t]} \rho(U(s), X_i(s) - u) - \rho(V(s), X_i(s) - u) W(du \times ds)\right|\Bigg).\end{aligned}$$

It follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}&\left(\frac{1}{n} \sum_{i=1}^{n} |A_i^U(t) - A_i^V(t)|\right)^2 \leqslant \\ &TC\left(\frac{1}{n} \sum_{i=1}^{n} \left(\Gamma_i^V(t) \vee \Gamma_i^U(t)\right)^2\right) \times \frac{1}{n} \sum_{i=1}^{n} \Bigg(\int_0^t |G(U(s), X_i(s), s) - G(V(s), X_i(s), s)|^2\, ds \\ &+ \int_0^t \left(\int_{\mathbb{R}^d} |\rho^2(U(s), X_i(s) - u) - \rho^2(V(s), X_i(s) - u)|\mu(du)\right)^2 ds \\ &+ \sup_{0\leqslant s\leqslant t} \left|\int_{\mathbb{R}^d \times (s,t]} \rho(U(s), X_i(s) - u) - \rho(V(s), X_i(s) - u) W(du \times ds)\right|^2\Bigg).\end{aligned}$$

Again appealing to Cauchy-Schwarz in the third line above, it follows that

$$\left(\frac{1}{n} \sum_{i=1}^{n} |A_i^U(t\wedge\eta_m) - A_i^V(t\wedge\eta_m)|\right)^2 \leqslant C\left(\frac{1}{n} \sum_{i=1}^{n} \left(\Gamma_i^V(t\wedge\eta_m) \vee \Gamma_i^U(t\wedge\eta_m)\right)^2\right)$$

$$\times \frac{1}{n}\sum_{i=1}^{n}\left(\int_0^{t\wedge\eta_m} d\left(U(s),V(s)\right)^2 ds\right.$$

$$\left.+\sup_{0\leqslant r\leqslant t\wedge\eta_m}\left|\int_{\mathbb{R}^d\times(r,t\wedge\eta_m]}\rho(U(s),X_i(s)-u)-\rho(V(s),X_i(s)-u)W(du\times ds)\right|^2\right).$$

Sending $n\to\infty$, we see that

$$d(\Phi U(t\wedge\eta_m),\Phi V(t\wedge\eta_m))^2\leqslant mC\int_0^t d(\Phi U(s\wedge\eta_m),\Phi V(s\wedge\eta_m))^2 ds$$

$$+mCE\Bigg[\sup_{0\leqslant r\leqslant t\wedge\eta_m}\left|\int_{\mathbb{R}^d\times(r,t\wedge\eta_m]}\rho(U(s),X_i(s)-u)-\rho(V(s),X_i(s)-u)W(du\times ds)\right|^2\Bigg|U,V,W\Bigg].$$

By additivity of the stochastic integral, we have (pathwise a.s.)

$$\sup_{0\leqslant r\leqslant t\wedge\eta_m}\left|\int_{\mathbb{R}^d\times(r,t\wedge\eta_m]}\rho(\Phi U(s),X_i(s)-u)-\rho(\Phi V(s),X_i(s)-u)W(du\times ds)\right|^2$$

$$\leqslant 4\sup_{0\leqslant r\leqslant t}\left|\int_{\mathbb{R}^d\times(0,r]}\rho(\Phi U(s\wedge\eta_m),X_i(s\wedge\eta_m)-u)-\rho(\Phi V(s\wedge\eta_m),X_i(s\wedge\eta_m)-u)W(du\times ds)\right|^2.$$

Applying this inequality and Doob's inequality, we then have

$$E\left[\sup_{0\leqslant r\leqslant t\wedge\eta_m}\left|\int_{\mathbb{R}^d\times(r,t\wedge\eta_m]}\rho(\Phi U(s),X_i(s)-u)-\rho(\Phi V(s),X_i(s)-u)W(du\times ds)\right|^2\right]$$

$$\leqslant 16\int_0^t E\left[\int_{\mathbb{R}^d}\left(\rho(\Phi U(s\wedge\eta_m),X_i(s\wedge\eta_m)-u)-\rho(\Phi V(s\wedge\eta_m),X_i(s\wedge\eta_m)-u)\right)^2\mu(du)\right]ds$$

$$\leqslant 16K\int_0^t E\left[d(\Phi U(s\wedge\eta_m),\Phi V(s\wedge\eta_m))^2\right]ds.$$

Taking expectations, we obtain for $t\leqslant T$,

$$E\left[d(\Phi U(t\wedge\eta_m),\Phi V(t\wedge\eta_m))^2\right]\leqslant mC\int_0^t E\left[d(\Phi U(s\wedge\eta_m),\Phi V(s\wedge\eta_m))^2\right]ds.$$

It follows that for any m and any $t\leqslant T$,

$$E\left[d(\Phi U(t\wedge\eta_m),\Phi V(t\wedge\eta_m))^2\right]=0.$$

Consequently, for each $m \in \mathbb{N}, t \in [0,T]$, $\Phi U(t \wedge \eta_m) = \Phi V(t \wedge \eta_m)$ almost surely. Taking $m \to \infty$, for each fixed t we have $\Phi U(t) = \Phi V(t)$ a.s. Applying Cauchy-Schwarz, we see that

$$\begin{aligned}
\big(E|A_i^U(t) - A_i^V(t)|\big)^2 \leqslant CE\left[\Gamma_i^U(t)^2 \vee \Gamma_i^V(t)^2\right] \times \\
E\Bigg[\bigg(\int_0^t |G(U(s), X_i(s), s) - G(V(s), X_i(s), s)|\, ds \\
+ \int_0^t \int_{\mathbb{R}^d} |\rho^2(U(s), X_i(s) - u) - \rho^2(V(s), X_i(s) - u)|\mu(du)ds \\
+ \sup_{0 \leqslant s \leqslant t} \left| \int_{\mathbb{R}^d \times (s,t]} \rho(U(s), X_i(s) - u) - \rho(V(s), X_i(s) - u) W(du \times ds) \right| \bigg)^2\Bigg] \\
\leqslant CE\left[\Gamma_i^U(t)^2 \vee \Gamma_i^V(t)^2\right] \int_0^t E\left[d(\Phi U(s), \Phi V(s))^2\right] ds = 0.
\end{aligned}$$

Consequently, for each $t \leqslant T$, $A_i^V(t) = A_i^U(t)$ a.s. Right continuity now implies that $A_i^U(\cdot) = A_i^V(\cdot)$ on $[0,T]$ a.s.. Consequently, $\Phi U(\cdot) = \Phi V(\cdot)$ on $[0,T]$ a.s.. □

3.3.5 Existence of a particle fixed point

We now turn to the existence of a consistent process U which satisfies $\Phi U = U$. We begin with Theorem 3.2.16, which shows that existence holds on some probability space. The results of [61] then allow us to combine this with Theorem 3.2.15 to prove Theorem 3.2.17.

We first introduce some notation for moduli of continuity ω on $C_{\mathbb{R}^d}[0,T]$ and ω' on $D_{\mathbb{R}}[0,T]$, which we define by

$$\omega(f, \delta, t) = \max_{0 \leqslant s \leqslant T} \max_{s \leqslant t \leqslant s+\delta} |f(t) - f(s)|,$$

$$\omega'(f, \delta, t) = \inf_{\{t_j\}} \max_{j} \sup_{s,t \in [t_j, t_{j+1})} |f(t) - f(s)|.$$

The infimum in the definition of ω' is taken over partitions of the form $0 < t_0 < t_1 < \cdots < t_{n-1} < T \leqslant t_n$ satisfying $\min_{0 \leqslant j \leqslant n} |t_j - t_{j-1}| > \delta$. Denote by $\bar{g}$ any continuous extension of g to $\overline{D}$.

Proof of Theorem 3.2.15. For $t \in [0, \frac{1}{n})$, we define the measure $V^{(n)}(t)$ by

$$V^{(n)}(t) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} h(X_i(0))\delta_{X_i(0)}.$$

We may then define for $t \in [0, \frac{1}{n}]$

$$\begin{aligned}
A_i^{(n)}(t) = \big[g(X_i(\tau_i(t))1_{\{\tau_i(t)>0\}} + h(X_i(0))1_{\{\tau_i(t)=0\}}\big] \exp\bigg\{ \int_{\tau_i(t)}^{t} G(V^{(n)}(s), X_i(s), s) \\
-\frac{1}{2}\int_{\mathbb{R}^d} \rho(V^{(n)}(s), X_i(s)-u)^2 \mu(du) ds + \int_{\mathbb{R}^d \times (\tau_i(t),t]} \rho(V^{(n)}(s-), X_i(s)-u) W(du \times ds)\bigg\} \\
+ \int_{\tau_i(t)}^{t} b_i(X_i(s)) \exp\bigg\{ \int_s^t G(V^{(n)}(r), X_i(r), r) - \frac{1}{2}\int_{\mathbb{R}^d} \rho(V^{(n)}(r), X_i(r), u)^2 \mu(du) dr \\
+ \int_{\mathbb{R}^d \times (s,t]} \rho(V^{(n)}(r-), X_i(r)-u) W(du \times dr)\bigg\} ds
\end{aligned} \tag{3.3.7}$$

which solves

$$\begin{aligned}
A_i^{(n)}(t) = g(X_i(\tau_i(t))1_{\{\tau_i(t)>0\}} + h(X_i(0))1_{\{\tau_i(t)=0\}} + \int_{\tau_i(t)}^{t} G(V^{(n)}(s), X_i(s), s) A_i^{(n)}(s) ds \\
+ \int_{\tau_i(t)}^{t} b_i(X_i(s)) ds + \int_{\mathbb{R}^d \times (\tau_i(t),t]} \rho(V^{(n)}(s-), X_i(s)-u) A_i^{(n)}(s-) W(du \times ds).
\end{aligned} \tag{3.3.8}$$

For $t \in [\frac{k}{n}, \frac{k+1}{n})$, we may recursively define $V^{(n)}$ by

$$V^{(n)}(t) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} A_i^{(n)}\left(\frac{k}{n}\right) \delta_{X_i(\frac{k}{n})} = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} A_i^{(n)}\left(\frac{\lfloor nt \rfloor}{n}\right) \delta_{X_i\left(\frac{\lfloor nt \rfloor}{n}\right)}$$

and $A_i^{(n)}(t)$ by (3.3.7) or equivalently (3.3.8). As before, by considering

$$\Xi^{(n)} = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} \delta_{A_i^{(n)}, X_i}$$

we may take all these almost sure limits to occur on a common set of full probability. Without loss of generality, $X_i(q) \notin \partial D$ for any $q \in \mathbb{Q}_+$ or $i \in \mathbb{N}$, so that for all $i, n \in \mathbb{N}$, $A_i^{(n)}(\cdot), \tau_i(\cdot) \in D_{\mathbb{R}_+}[0,\infty)$ will be continuous at $\frac{k}{n}$ for all $i, k, n \in \mathbb{N}$. By construction, $V^{(n)}(\cdot) \in D_{\mathcal{M}_+(D)}[0,\infty)$,

$V^{(n)}$ is adapted to $\{\mathcal{F}_t^{W,\{X_i\}}\}_{t\geqslant 0}$, and the family $\{(W, X_i, V^{(n)})\}_{i=1}^{\infty}$ is exchangeable for each n. In particular, $V^{(n)}(\cdot)$ is consistent. Define $Z_i^{(n)}(t)$ by

$$\begin{aligned} Z_i^{(n)}(t) = \bar{g}(X_i(t)) + \int_0^t G(V^{(n)}(s), X_i(s), s) A_i^{(n)}(s) ds + \int_0^t b(X_i(s)) ds \\ + \int_{\mathbb{R}^d \times (0,t]} \rho(V^{(n)}(s-), X_i(s) - u) A_i^{(n)}(s-) W(du \times ds). \end{aligned}$$

We note that $\{Z_i^{(n)}(\cdot)\}_n$ is tight. This can be seen by applying Kolmogorov-Censov to the terms which depend on n. With this notation, we have

$$A_i^{(n)}(t) = \bar{g}(X_i(t)) + Z_i^{(n)}(t) - Z_i^{(n)}(\tau_i(t)) + (h(X_i(0)) - \bar{g}(X_i(0)) 1_{\{\tau_i(t)=0\}}. \tag{3.3.9}$$

We will control the modulus of continuity of $A_i^{(n)}$ pathwise by constructing an appropriate partition $\mathcal{P}$ of $[0, T+\delta]$ for sufficiently small δ.

Define $\gamma_i = \inf\{t > 0 : X_i(t) \in \partial D\}$ and take any $\delta_0 > 0$ satisfying $\delta_0 < \gamma_i$. If $\gamma_i < T$, then add γ_i to $\mathcal{P}$. $\tau_i(\cdot)$ is right continuous with left limits, so on $[0, T+\delta_0]$ there are only finitely many points t with $\tau_i(t) - \tau_i(t-) \geqslant \delta_0$. Add these points to $\mathcal{P}$ and let $\delta > 0$ be such that if $\tau_i(t_1) - \tau_i(t_1-) \geqslant \delta_0$ and $\tau_i(t_2) - \tau_i(t_2-) \geqslant \delta_0$ then $|t_1 - t_2| > 2\delta$ and $2\delta < \delta_0$.

Keeping all the previously added points, refine $\mathcal{P}$ to a partition $\{t_j\}$ satisfying $\delta < t_j - t_{j-1} < 4\delta$. It will suffice to focus on controlling the modulus of continuity of $Z_i^{(n)}(\tau_i(t))$ on this partition.

Fix j and take any $s, t \in [t_{j-1}, t_j)$ with $s < t$. Note that $\tau_i(s) = \tau_i(t)$ unless there exists $u \in (s, t]$ with $X_i(u) \in \partial D$. Call $\gamma_i^s = \inf\{u \geqslant s : X_i(u) \in \partial D\}$. If $\gamma_i^s = s$ then $s = \tau_i(s) \leqslant \tau_i(t) \leqslant t$ and $|\tau_i(t) - \tau_i(s)| < 4\delta < 2\delta_0$. Suppose that $s < \gamma_i^s \leqslant t$. Since $X_i(\gamma_i^s) \in \partial D, \tau_i(\gamma_i^s) = \gamma_i^s$ and we have $s < \tau_i(\gamma_i^s) \leqslant \tau_i(t) \leqslant t$. Note that $|\tau_i(t) - \tau_i(\gamma_i^s)| \leqslant t_j - t_{j-1} < 4\delta$. By construction of $\mathcal{P}$, we also have $|\tau_i(\gamma_i^s) - \tau_i(\gamma_i^s-)| < \delta_0$ and by definition of γ_i^s, we have $\tau_i(\gamma_i^s-) = \tau_i(s)$. It then follows that for any j and any $s < t$ with $s, t \in [t_{j-1}, t_j)$,

$$|Z_i^{(n)}(\tau_i(t)) - Z_i^{(n)}(\tau_i(s))| \leqslant |Z_i^{(n)}(\tau_i(t)) - Z_i^{(n)}(\tau_i(\gamma_i^s))| + |Z_i^{(n)}(\tau_i(\gamma_i^s)) - Z_i^{(n)}(\tau_i(\gamma_i^s-))|$$

$$< 2\omega(Z_i^{(n)}, 2\delta_0, T + \delta_0).$$

It follows that $\omega'(Z^{(n)}, \delta_0, T) \leqslant 2w(Z_i^{(n)}, 2\delta_0, T + \delta_0)$. Noting that the last term in (3.3.9) is constant on $[0, \gamma_i)$ and $[\gamma_i, \infty)$, we see that

$$\omega'(A_i^{(n)}, \delta, T) \leqslant \omega(\bar{g} \circ X_i, \delta, T + \delta) + 3\omega(Z_i^{(n)}, 2\delta_0, T + \delta_0).$$

For each i, the tightness of $\{A_i^{(n)}\}_n$ in $D_{\mathbb{R}_+}[0, \infty)$ now follows from the tightness of $\{Z_i^{(n)}\}_n$ in $C_{\mathbb{R}^d}[0, \infty)$; see [33, Theorem 3.7.2]. We have $P(\exists s : X_i(s) \in \partial D \text{ and } X_j(s) \in \partial D) = 0$ for $i \neq j$. For all n, $A_i^{(n)}$ is continuous off of the set $\{s : X_i(s) \in \partial D\}$. It follows that for any k, any distributional limit point of $(A_1^{(n)}, \ldots, A_k^{(n)})$ will have components with no simultaneous discontinuities. $\{A_i^{(n)}\}_{i \leqslant k}$ is tight in $D_{\mathbb{R}_+}[0, \infty)^k$, so it follows that $\{A_i^{(n)}\}_{i \leqslant k}$ is tight in $D_{\mathbb{R}_+^k}[0, \infty)$. This implies tightness of $\{A_i^{(n)}\}_{i=1}^{\infty}$ in $D_{\mathbb{R}_+^{\infty}}[0, \infty)$. X_i is continuous, so this implies tightness of $(\{A_i^{(n)}\}, \{X_i\})$ in $D_{(\mathbb{R}_+ \times \overline{D})^{\infty}}[0, \infty)$. With the notation

$$\Xi^{(n)} = \lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} \delta_{A_i^{(n)}, X_i}, \qquad \Xi^{(n)}(t) = \lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} \delta_{A_i^{(n)}(t), X_i(t)}$$

[58, Lemma 4.4] implies tightness of $(\{A_i^{(n)}(\cdot)\}, \{X_i(\cdot)\}, \Xi^{(n)}(\cdot))$ in $D_{(\mathbb{R} \times \overline{D})^{\infty} \times \mathcal{P}(\mathbb{R} \times \overline{D}))}[0, \infty)$. Moreover, [58, Lemma 4.4] shows that any subsequential weak limit point $(\{A_i(\cdot)\}, \{X_i(\cdot)\}, \Xi(\cdot))$ of $(\{A_i^{(n)}(\cdot)\}, \{X_i(\cdot)\}, \Xi^{(n)}(\cdot))$ has

$$\Xi(t) = \lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} \delta_{A_i(t), X_i(t)}.$$

By [58, Lemma 4.6], any such limit point will also have $\Xi(\cdot) \in C_{\mathcal{P}(\mathbb{R} \times \overline{D})}[0, \infty)$. In the arguments that follow, we will send $n \to \infty$ along a subsequence for which weak convergence holds. For notational convenience, we will not keep track of this subsequence. Let $\eta_n(t) = n^{-1} \lfloor nt \rfloor$; it then follows from [33, Proposition 3.6.5] that for such a limit point $(\{A_i^{(n)}(\cdot)\}, \{X_i(\cdot)\}, \Xi^{(n)} \circ \eta_n(\cdot)) \Longrightarrow (\{A_i(\cdot)\}, \{X_i(\cdot)\}, \Xi(\cdot))$ in $D_{(\mathbb{R}_+ \times \overline{D})^{\infty} \times \mathcal{P}(\mathbb{R} \times \overline{D})}[0, \infty)$.

Fix $\alpha > 0$ and let $V^{(n),\alpha}(t)$ denote the positive measure determined by the map on $C_c(\mathbb{R}^d)$

$$\varphi \mapsto \int_{\mathbb{R} \times \overline{D}} (a \wedge \alpha) \varphi(x) \Xi^{(n)} \circ \eta_n(t)(da \times dx)$$

and let $V^{\alpha}(t)$ be given by the same expression with $\Xi^{(n)} \circ \eta_n$ replaced by Ξ. It follows immediately that $(\{A_i^{(n)}(\cdot)\}, \{X_i(\cdot)\}, V^{(n),\alpha} \circ \eta_n(\cdot))$ is tight in $D_{(\mathbb{R}_+ \times \overline{D})^\infty \times \mathcal{M}_+(D)}[0,\infty)$. Lemma 3.2.6 gives the moment bound needed to show convergence of $V^{(n),\alpha}(\cdot) \to V^{(n)}(\cdot)$ as $\alpha \to \infty$, which gives tightness of $(\{A_i^{(n)}\}, \{X_i\}, V^{(n)})$. As above, any limit point $(\{A_i\}, X_i, V)$ in $D_{(\mathbb{R}_+ \times \overline{D})^\infty \times \mathcal{M}_+(D)}[0,\infty)$ satisfies

$$V(t) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} A_i(t)\delta_{X_i(t)}.$$

Recall that in the coupling we have constructed, $A_i^{(n)}$ is given by (3.3.8). Using continuity of ρ and G and [62, Proposition 7.4], any limit point $(\{A_i\}, \{X_i\}, V, W)$ of $(\{A_i^{(n)}\}, \{X_i\}, V^{(n)}, W)$ in $D_{(\mathbb{R}_+ \times \overline{D})^\infty \times \mathcal{M}_+(D) \times \mathcal{H}^{-1}}[0,\infty)$ will have A given by (3.3.7) with $A_i^{(n)}, V^{(n)}$ replaced by (A_i, V).

Finally, note that in the coupling we have constructed, $V^{(n)}$ is consistent with $(\{X_i\}, W)$ for each n. Exchangeability of $\{(W, X_i, V^{(n)})\}_{i=1}^{\infty}$ is preserved under joint distributional limits of $(V^{(n)}, W, \{X_i\})$ and so for any weak limit point the family $\{(W, X_i, V)\}_{i=1}^{\infty}$ will be exchangeable. The remainder of the definition of compatibility for any limit point $(V, W, \{X_i\})$ then follows from [61, Lemma 3.5] (we are in the setting of the last sentence of [61, Remark 3.6]). $\square$

Proof of Theorem 3.2.17. Strong existence follows from weak existence (Theorem 3.2.16) and pathwise uniqueness of jointly compatible solutions (Theorem 3.2.15); see [61, Lemma 2.10, Theorem 1.5]. In particular, it follows that there exists a Borel measurable function with $G(\cdot,\cdot) : C_{\overline{D}^\infty \times \mathcal{H}^{-1}}[0,\infty) \to C_{\mathcal{M}_+(D)}[0,\infty)$ with the property that $G(\{X_i\}, W)$ satisfies $\Phi G(\{X_i\}, W) = G(\{X_i\}, W)$ and that the process $G(\{X_i\}, W)(\cdot)$ is $\{\mathcal{F}_t^{\{X_i\},W}\}_{t\geqslant 0}$ adapted [61, Proposition 2.13]. Such a process is measurable with respect to the tail sigma algebra of the exchangeable sequence $\{(X_i, W)\}_{i=1}^{\infty}$ and therefore the vector $(G(\{X_i\}, W), W)$ is independent of the family $\{X_i\}$. It follows that there is a Borel function $F : C_{\mathcal{H}^{-1}}[0,\infty) \to C_{\mathcal{M}_+(D)}[0,\infty)$ with $F(W) = G(\{X_i\}, W)$ almost surely and moreover $\{\mathcal{F}_t^{\{X_i\},W}\}_{t\geqslant 0}$ adaptedness of $G(\{X_i\}, W)$ implies that $F(W)(\cdot)$ is $\{\mathcal{F}_t^{W}\}_{t\geqslant 0}$ adapted. $\square$

Bibliography

[1] T. Alberts, K. Khanin, and J. Quastel, *The continuum directed random polymer*, Journal of Statistical Physics, 154 (2013), pp. 305–326.

[2] ——, *The intermediate disorder regime for directed polymers in dimension* $1+1$, Ann. Probab., 42 (2014), pp. 1212–1256.

[3] G. Amir, I. Corwin, and J. Quastel, *Probability distribution of the free energy of the continuum directed random polymer in* $1+1$ *dimensions*, Comm. Pure Appl. Math., 64 (2011), pp. 466–537.

[4] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, vol. 118 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2010.

[5] G. Aubrun, *A sharp small deviation inequality for the largest eigenvalue of a random matrix*, in Séminaire de Probabilités XXXVIII, vol. 1857 of Lecture Notes in Math., Springer, Berlin, 2005, pp. 320–337.

[6] M. Balázs, E. Cator, and T. Seppäläinen, *Cube root fluctuations for the corner growth model associated to the exclusion process*, Electron. J. Probab., 11 (2006), pp. no. 42, 1094–1132 (electronic).

[7] Y. Baryshnikov, *GUEs and queues*, Probab. Theory Related Fields, 119 (2001), pp. 256–274.

[8] I. Ben-Ari, *Large deviations for partition functions of directed polymers in an IID field*, Ann. Inst. Henri Poincaré Probab. Stat., 45 (2009), pp. 770–792.

[9] G. Ben Arous, A. Dembo, and A. Guionnet, *Aging of spherical spin glasses*, Probab. Theory Related Fields, 120 (2001), pp. 1–67.

[10] G. Ben Arous and A. Guionnet, *Large deviations for Wigner's law and Voiculescu's non-commutative entropy*, Probab. Theory Related Fields, 108 (1997), pp. 517–542.

[11] L. Bertini and N. Cancrini, *The stochastic heat equation: Feynman-kac formula and intermittence*, Journal of Statistical Physics, 78 (1995), pp. 1377–1401.

[12] G. Biroli, J.-P. Bouchaud, and M. Potter, *Extreme value problems in random matrix theory and other disordered systems*, J. Stat. Mech., (2007), p. P07019.

[13] V. I. Bogachev, *Measure theory. Vol. I, II*, Springer-Verlag, Berlin, 2007.

[14] E. Bolthausen, *A note on the diffusion of directed polymers in a random environment*, Comm. Math. Phys., 123 (1989), pp. 529–534.

[15] A. Borodin and I. Corwin, *Macdonald processes*, Probab. Theory Related Fields, 158 (2014), pp. 225–400.

[16] ——, *Moments and Lyapunov exponents for the parabolic Anderson model*, Ann. Appl. Probab., 24 (2014), pp. 1172–1198.

[17] A. Borodin, I. Corwin, and P. Ferrari, *Free energy fluctuations for directed polymers in random media in* $1+1$ *dimension*, Comm. Pure Appl. Math., 67 (2014), pp. 1129–1214.

[18] A. Borodin and S. Péché, *Airy kernel with two sets of parameters in directed percolation and random matrix theory*, J. Stat. Phys., 132 (2008), pp. 275–290.

[19] R. Carmona, F. Delarue, and A. Lachapelle, *Control of McKean-Vlasov dynamics versus mean field games*, Math. Financ. Econ., 7 (2013), pp. 131–166.

[20] R. A. Carmona and S. A. Molchanov, *Parabolic Anderson problem and intermittency*, Mem. Amer. Math. Soc., 108 (1994), pp. viii+125.

[21] X. Chen, *Precise intermittency for the parabolic Anderson equation with an* (1 + 1)*-dimensional time-space white noise*, Ann. Inst. Henri Poincaré Probab. Stat., 51 (2015), pp. 1486–1499.

[22] F. Comets, N. Gantert, and O. Zeitouni, *Quenched, annealed and functional large deviations for one-dimensional random walk in random environment*, Probab. Theory Related Fields, 118 (2000), pp. 65–114.

[23] F. Comets, T. Shiga, and N. Yoshida, *Probabilistic analysis of directed polymers in a random environment: a review*, in Stochastic analysis on large scale interacting systems, vol. 39 of Adv. Stud. Pure Math., Math. Soc. Japan, Tokyo, 2004, pp. 115–142.

[24] D. Conus, M. Joseph, and D. Khoshnevisan, *On the chaotic character of the stochastic heat equation, before the onset of intermitttency*, Ann. Probab., 41 (2013), pp. 2225–2260.

[25] I. Corwin, *The Kardar-Parisi-Zhang equation and universality class*, Random Matrices Theory Appl., 1 (2012), pp. 1130001, 76.

[26] I. Corwin, J. Quastel, and D. Remenik, *Renormalization fixed point of the KPZ universality class*, J. Stat. Phys., 160 (2015), pp. 815–834.

[27] M. Cranston, D. Gauthier, and T. S. Mountford, *On large deviation regimes for random media models*, Ann. Appl. Probab., 19 (2009), pp. 826–862.

[28] D. Crisan, C. Janjigian, and T. G. Kurtz, *Particle representations for stochastic partial differential equations with boundary conditions*, (2016).

[29] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, vol. 38 of Stochastic Modelling and Applied Probability, Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[30] F. den Hollander, *Random polymers*, vol. 1974 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2009. Lectures from the 37th Probability Summer School held in Saint-Flour, 2007.

[31] E. Emrah, *Limit shapes for inhomogeneous corner growth models with exponential and geometric weights*, Electron. Comm. Probab., 21 (2016), pp. 1–16.

[32] E. Emrah and C. Janjigian, *Large deviations for some corner growth models with inhomogeneity*, (2015).

[33] S. N. Ethier and T. G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[34] P. J. Forrester and E. M. Rains, *Interpretations of some parameter dependent generalizations of classical matrix ensembles*, Probab. Theory Related Fields, 131 (2005), pp. 1–61.

[35] D. Forster, D. R. Nelson, and M. J. Stephen, *Large-distance and long-time properties of a randomly stirred fluid*, Phys. Rev. A, 16 (1977), pp. 732–749.

[36] T. Funaki, *A certain class of diffusion processes associated with nonlinear parabolic equations*, Z. Wahrsch. Verw. Gebiete, 67 (1984), pp. 331–348.

[37] N. Georgiou and T. Seppäläinen, *Large deviation rate functions for the partition function in a log-gamma distributed random potential*, Ann. Probab., 41 (2013), pp. 4248–4286.

[38] J. Gravner, C. A. Tracy, and H. Widom, *Limit theorems for height fluctuations in a class of discrete space and time growth models*, J. Statist. Phys., 102 (2001), pp. 1085–1132.

[39] M. Gubinelli and N. Perkowski, *Kpz reloaded*, (2015).

[40] M. Hairer, *Solving the KPZ equation*, Ann. of Math. (2), 178 (2013), pp. 559–664.

[41] ——, *A theory of regularity structures*, Invent. Math., 198 (2014), pp. 269–504.

[42] T. Halpin-Healy and K. A. Takeuchi, *A KPZ cocktail—shaken, not stirred . . . toasting 30 years of kinetically roughened surfaces*, J. Stat. Phys., 160 (2015), pp. 794–814.

[43] T. Halpin-Healy and Y.-C. Zhang, *Kinetic roughening phenomena, stochastic growth, directed polymers and all that. aspects of multidisciplinary statistical mechanics*, Physics Reports, 254 (1995), pp. 215 – 414.

[44] B. M. Hambly, J. B. Martin, and N. O'Connell, *Concentration results for a Brownian directed percolation problem*, Stochastic Process. Appl., 102 (2002), pp. 207–220.

[45] J. M. Harrison and R. J. Williams, *On the quasireversibility of a multiclass Brownian service station*, Ann. Probab., 18 (1990), pp. 1249–1268.

[46] D. A. Huse and C. L. Henley, *Pinning and roughening of domain walls in Ising systems due to random impurities*, Phys. Rev. Lett., 54 (1985), pp. 2708–2711.

[47] D. A. Huse, C. L. Henley, and D. S. Fisher, *Huse, henley, and fisher respond*, Phys. Rev. Lett., 55 (1985), pp. 2924–2924.

[48] J. Z. Imbrie and T. Spencer, *Diffusion of directed polymers in a random environment*, J. Statist. Phys., 52 (1988), pp. 609–626.

[49] C. Janjigian, *Large deviations of the free energy in the O'Connell-Yor polymer*, J. Stat. Phys., 160 (2015), pp. 1054–1080.

[50] ——, *Right tail large deviations in Brownian directed percolation*, (2016).

[51] K. Johansson, *Shape fluctuations and random matrices*, Comm. Math. Phys., 209 (2000), pp. 437–476.

[52] K. Johansson, *Random growth and random matrices*, in European Congress of Mathematics, C. Casacuberta, M.-R. R.M, J. Verdera, and X.-D. S., eds., Springer, 2001, pp. 445–456.

[53] O. Kallenberg, *Foundations of modern probability*, Probability and its Applications (New York), Springer-Verlag, New York, second ed., 2002.

[54] M. Kardar, *Roughening by impurities at finite temperatures*, Phys. Rev. Lett., 55 (1985), p. 2923.

[55] H. Kesten, *Aspects of first passage percolation*, in École d'été de probabilités de Saint-Flour, XIV—1984, vol. 1180 of Lecture Notes in Math., Springer, Berlin, 1986, pp. 125–264.

[56] D. Khoshnevisan and K. Kim, *Nonlinear noise excitation of intermittent stochastic PDEs and the topology of LCA groups*, Ann. Probab., 43 (2015), pp. 1944–1991.

[57] V. N. Kolokoltsov, *Nonlinear Markov processes and kinetic equations*, vol. 182 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2010.

[58] P. M. Kotelenez and T. G. Kurtz, *Macroscopic limits for stochastic partial differential equations of McKean-Vlasov type*, Probab. Theory Related Fields, 146 (2010), pp. 189–222.

[59] J. Krug and H. Spohn, *Kinetic roughening of growing surfaces*, in Solids far from Equilibrium., C. Godreche, ed., Aléa-Saclay, Monographs and Texts in Statistical Physics 1, Cambridge Univ. Press, New York, 1992, p. 412?525.

[60] M. Kuczma, *An introduction to the theory of functional equations and inequalities*, Birkhäuser Verlag, Basel, second ed., 2009. Edited and with a preface by Attila Gilányi.

[61] T. G. Kurtz, *Weak and strong solutions of general stochastic models*, Electron. Commun. Probab., 19 (2014), pp. no. 58, 16.

[62] T. G. Kurtz and P. E. Protter, *Weak convergence of stochastic integrals and differential equations. II. Infinite-dimensional case*, in Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995), vol. 1627 of Lecture Notes in Math., Springer, Berlin, 1996, pp. 197–285.

[63] T. G. Kurtz and J. Xiong, *Particle representations for a class of nonlinear SPDEs*, Stochastic Process. Appl., 83 (1999), pp. 103–126.

[64] M. Ledoux, *Deviation inequalities on largest eigenvalues*, in Geometric aspects of functional analysis, vol. 1910 of Lecture Notes in Math., Springer, Berlin, 2007, pp. 167–219.

[65] M. Ledoux and B. Rider, *Small deviations for beta ensembles*, Electron. J. Probab., 15 (2010), pp. 1319–1343.

[66] T. M. Liggett, *An improved subadditive ergodic theorem*, Ann. Probab., 13 (1985), pp. 1279–1285.

[67] H. Matsumoto and M. Yor, *A relationship between Brownian motions with opposite drifts via certain enlargements of the Brownian filtration*, Osaka J. Math., 38 (2001), pp. 383–398.

[68] H. P. McKean, Jr., *Propagation of chaos for a class of non-linear parabolic equations.*, in Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967), Air Force Office Sci. Res., Arlington, Va., 1967, pp. 41–57.

[69] J. Moriarty and N. O'Connell, *On the free energy of a directed polymer in a Brownian environment*, Markov Process. Related Fields, 13 (2007), pp. 251–266.

[70] N. O'Connell, *Directed polymers and the quantum Toda lattice*, Ann. Probab., 40 (2012), pp. 437–458.

[71] N. O'Connell and M. Yor, *Brownian analogues of Burke's theorem*, Stochastic Process. Appl., 96 (2001), pp. 285–304.

[72] ——, *A representation for non-colliding random walks*, Electron. Comm. Probab., 7 (2002), pp. 1–12 (electronic).

[73] K. Oelschläger, *A martingale approach to the law of large numbers for weakly interacting stochastic processes*, Ann. Probab., 12 (1984), pp. 458–479.

[74] A. Okounkov, *Infinite wedge and random partitions*, Selecta Math. (N.S.), 7 (2001), pp. 57–81.

[75] J. Quastel, *Introduction to KPZ*, in Current developments in mathematics, 2011, Int. Press, Somerville, MA, 2012, pp. 125–194.

[76] J. Quastel and H. Spohn, *The one-dimensional KPZ equation and its universality class*, J. Stat. Phys., 160 (2015), pp. 965–984.

[77] F. Rassoul-Agha and T. Seppäläinen, *A course on large deviations with an introduction to Gibbs measures*, vol. 162 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2015.

[78] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, vol. 293 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, third ed., 1999.

[79] R. T. Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970.

[80] T. Seppäläinen, *Coupling the totally asymmetric simple exclusion process with a moving interface*, Markov Process. Related Fields, 4 (1998), pp. 593–628. I Brazilian School in Probability (Rio de Janeiro, 1997).

[81] ——, *Large deviations for increasing sequences on the plane*, Probab. Theory Related Fields, 112 (1998), pp. 221–244.

[82] ——, *Lecture notes on the corner growth model.* `http://www.math.wisc.edu/~seppalai/cornergrowth-book/ajo.pdf`, 2009.

[83] T. Seppäläinen and B. Valkó, *Bounds for scaling exponents for a* $1+1$ *dimensional directed polymer in a Brownian environment*, ALEA Lat. Am. J. Probab. Math. Stat., 7 (2010), pp. 451–476.

[84] M. Sion, *On general minimax theorems*, Pacific J. Math., 8 (1958), pp. 171–176.

[85] H. Spohn, *KPZ scaling theory and the semidiscrete directed polymer model*, in Random matrix theory, interacting particle systems, and integrable systems, vol. 65 of Math. Sci. Res. Inst. Publ., Cambridge Univ. Press, New York, 2014, pp. 483–493.

[86] D. W. Stroock and S. R. S. Varadhan, *Diffusion processes with boundary conditions*, Comm. Pure Appl. Math., 24 (1971), pp. 147–225.

[87] C. A. Tracy and H. Widom, *Level-spacing distributions and the Airy kernel*, Comm. Math. Phys., 159 (1994), pp. 151–174.

[88] H. van Beijeren, R. Kutner, and H. Spohn, *Excess noise for driven diffusive systems*, Phys. Rev. Lett., 54 (1985), pp. 2026–2029.

[89] V. Vargas, *Strong localization and macroscopic atoms for directed polymers*, Probab. Theory Related Fields, 138 (2007), pp. 391–410.

[90] J. B. Walsh, *An introduction to stochastic partial differential equations*, in École d'été de probabilités de Saint-Flour, XIV—1984, vol. 1180 of Lecture Notes in Math., Springer, Berlin, 1986, pp. 265–439.